

Renzo Shamey
Editor

Encyclopedia of Color Science and Technology

Encyclopedia of Color Science and Technology

Renzo Shamey
Editor

Encyclopedia of Color Science and Technology

Second Edition

With 1693 Figures and 138 Tables

 Springer

Editor

Renzo Shamey
North Carolina State University
Color Science & Imaging Laboratories
Raleigh, NC, USA

ISBN 978-3-030-89861-8 ISBN 978-3-030-89862-5 (eBook)
<https://doi.org/10.1007/978-3-030-89862-5>

1st edition: © Springer Science+Business Media New York 2016

© Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

I would like to dedicate this work to my family for their loving support throughout my life, my son Sean Araz Luca, my wife Jiajun, and my entire research team who have helped me with my discoveries in the fascinating world of color science and technology.

Preface

It is with great pleasure that I present to you the second edition of the *Encyclopedia of Color Science and Technology*, a comprehensive and groundbreaking resource that delves into the intricate world of color. This project brings together the collective wisdom of experts in the fields of color science, technology, physiology, and psychology, providing a valuable reference for researchers, practitioners, and enthusiasts.

Color, as a phenomenon, has captivated our imaginations since time immemorial. Its influence spans a myriad of disciplines, from art and design to physics, biology, and beyond. This new edition of the encyclopedia continues to explore the multifaceted nature of color, examining its scientific underpinnings, technological applications, physiological mechanisms, and psychological implications. Our contributors provide detailed explanations and insights into the principles and techniques that underpin color science. Moreover, the encyclopedia explores the practical applications of color, such as color in imaging, displays, printing, and various industries, demonstrating the profound impact that color technology has on our daily lives.

Understanding color goes beyond its physical and technological aspects. The physiology of color perception and the intricate workings of the human visual system play a fundamental role in how we experience and interpret color. The fascinating realms of color vision, and the neural processes that enable us to perceive the rich tapestry of colors that surround us are among topics examined. Additionally, the psychology of color including the emotional, cognitive, and cultural dimensions of color, as well as its effects on human behavior, perception, and communication are explored. With over 300 new and revised entries, this edition serves as an extension of the authoritative compendium of knowledge in the ever-evolving realm of color science and technology.

I would like to express my heartfelt gratitude to the esteemed section editors who have worked tirelessly to shape this encyclopedia into a definitive resource. Their expertise, dedication, and commitment to excellence have been instrumental in ensuring the accuracy and comprehensiveness of each entry. Their vision and guidance have played a pivotal role in providing entries with unmatched depth and breadth in the domain.

I would especially like to thank Professor Rolf Kuehni for his exceptional leadership as the section editor of two vision science topics. Professor Kuehni's profound knowledge and extensive research in the field of color and his guidance and meticulous attention to detail have been instrumental in ensuring the scientific rigor and accuracy of the entries within this work.

I would also like to extend my heartfelt thanks to Professor Harold S. Freeman, the section editor for materials. Professor Freeman's expertise and profound understanding of colorants and coloration technology have been indispensable in exploring the practical applications and advancements in this arena. His insightful contributions have enriched this encyclopedia, providing invaluable insights into colorants and their various applications.

Furthermore, I would like to express my sincere appreciation to Professor Caivano, the section editor for art and design. Professor Caivano's expertise has greatly enhanced the understanding of how we perceive and experience the artistic aspects of color. Lastly, I would like to extend my deepest gratitude to Professors van Bommel, the section editor for lighting; Professor MacDonald, the section editor for language and categorization; and Dr. Reinhard, the section editor for computer graphics. Their expertise and guidance have helped shape this encyclopedia. Their commitment to excellence and their unwavering dedication to their respective sections have left an indelible mark on this project.

I express my deepest appreciation to all other section editors for their valuable time and effort, contributors, and the entire editorial and production teams who have played a vital role in bringing the *Encyclopedia of Color Science and Technology* to fruition. Their meticulous attention to detail and unwavering commitment to quality have been invaluable in creating a volume that is not only informative but also aesthetically pleasing.

Before concluding, I would like to take a moment to express a personal note. The COVID-19 pandemic has had a profound impact on the world, disrupting various aspects of our lives, including academic and publishing endeavors. The challenges posed by the pandemic, such as restrictions on travel, limited access to research facilities, and remote working arrangements, have unavoidably affected the timeline of this project. Additionally, as the Editor-in-Chief, I faced unexpected health challenges that slowed down the finalization of this volume. I am immensely grateful for the support and understanding extended to me during this time, both from the editorial team and the contributors. Their unwavering dedication and assistance have been invaluable in navigating these unforeseen circumstances.

It is my hope that the *Encyclopedia of Color Science and Technology* will serve as a cornerstone for researchers, educators, and practitioners, inspiring further exploration and innovation in the captivating realm of color. May it illuminate your path and deepen your understanding of the vibrant world of color.

Raleigh, USA
July 2023

Renzo Shamey

List of Topics

Art and Design

Section Editor: *Prof. José Luis Caivano*

Appearance
Art and Fashion Color Design
Chevreul, Michel-Eugène
Color Circle
Color Combination
Color Contrast
Color Field Painting
Color Harmony
Color in Hospitals
Color in Schools and Offices
Color in Transport
Color Order Systems
Color Pollution
Color Psychology
Color Scheme
Color Trends
Complementary Colors
Environmental Color Design
Functionality of Color
Impressionism
Interactions of Color and Light
Mosaics
Neo-impressionism
Optical Art
Palette
Pastel Colors
Philosophy of Color
Polychromy
Primary Colors
Psychological Color Effects
RYB Color

Tincture
Traditional Color Theory

Color Engineering: Computations

Section Editor: *Gaurav Sharma*

Chromatic Contrast Sensitivity
Clapper-Yule Model
Demichel Equations
Gamut Volume
Vora Value

Color Science: Computations

Section Editor: *Prof. Stephen Westland*

Interpolation of Spectral Data
Transformation of Color Matching Functions

Computer Graphics

Section Editor: *Dr. Erik Reinhard*

Color Transfer
Compositing and Chroma Keying
Global Illumination
High Dynamic Range Imaging
Image Quality
Intrinsic Imaging
Light Stage
Non-photorealistic Rendering
Rendering Natural Phenomena
Sampling Problems in Computer Graphics

Sampling, Reconstruction, Aliasing, and
Anti-aliasing
Stereo and Anaglyph Images
Texture Measurement, Modeling, and Computer
Graphics
Transient Imaging

Language and Categorisation

Section Editor: *Dr. Lindsay MacDonald*

Ancient Color Terminology
Bayesian Approaches to Color Category
Learning
Berlin and Kay Theory
Bilingual Color Naming/Categories
Chinese Color Language
Color Categorical Perception
Color Categories and Vagueness
Color Categorization and Naming in Inherited
Color Vision Deficiencies
Color Categorization: Patterns and Mechanisms
of Evolution
Color Category Learning in Naming-Game
Simulations
Color Constancy and Categories
Color Dictionaries and Corpora
Color Name Applications in Computer Vision
Color Preference
Cross-cultural Communication
Crowdsourcing of Multilingual Color Names
Dynamics of Color Category Formation and
Boundaries
Effects of Color Terms on Color Perception and
Cognition
Evolution of Color Expression
Infant Color Categories
Materials and the Origins of Abstract Color
Naming
Mechanisms of Color Categorization
Memory Color
Mesoamerican Color Survey Digital Archive
Psychological Color Space and Color Terms
Psychologically Pure Colors
Stroop Effect
Translation of Color Language
Vantage Theory of Color
World Color Survey

Lighting

Section Editor: *Prof. Wout van Bommel*

Architectural Lighting
Automotive Front Lighting System (Status 2020)
Autonomous Vehicle (AV) Lighting Systems
Blacklight Lamp
Carbon Arc Lamp
Coefficient of Utilization, Lumen Method
Combustion Lamp
Daylight Factor
Daylighting
Diffuse Lighting
Electrical Control Gear for Lamps
Emergency Lighting
Halogen Lamp
High Pressure Mercury Lamp
High- and Low-Pressure Sodium Lamp
Human-Centric Lighting
Incandescence
Incandescent Lamp
Induction Lamp
Interior Lighting
Internet of Things and LEDs
Kruithof Curve
Light Distribution
Light Pollution
Light-Emitting Diode, LED
Light-Emitting Diode, OLED
Lighting Controls
Lighting for Visual Satisfaction
Lumen Depreciation
Luminaires
Melanopic Equivalent Daylight Illuminance
Metal Halide Lamp
Metamerism
Neon Lamp
Nonvisual Lighting Effects and Their Impact
on Health and Well-Being
Phosphors and Fluorescent Powders
Quantum Dots
Road Lighting
Sleep, Alertness, and Light
Sports Lighting
Tubular and Compact Fluorescent Lamp
Visible Light Communication
Vision and Age
Xenon Lamp

Materials: General**Section Editor: Prof. Dr. Harold S. Freeman**

Alizarin
 Batik
 Chromophore
 Colorant, Environmental Aspects
 Colorant, Natural
 Colorant, Nonlinear Optical
 Colorant, Photochromic
 Colorant, Textile
 Colorant, Thermochromic
 Coloration, Fastness
 Coloration, Hair
 Coloration, Ink-Jet Printing
 Coloration, Mordants
 Coloration, Textile
 Colour Index
 Dye
 Dye, DSSC
 Dye, Functional
 Dye, Ink-Jet
 Dye, Metal Complex
 Dye: Photodynamic Therapy
 Ferroelectric Liquid Crystal Cells for Modern
 Applications in Displays and Photonics
 Finish, Textile
 Inorganic, Hybrid, and Functional Pigments
 Mauveine
 Melanin
 Pigment, Ceramic
 Pigments, Organic

Metrology: Concepts and Devices**Section Editor: Dr. Tzeng-Yow Lin**

Gloss Meter
 Goniophotometer
 Illuminance Meter
 Instrument: Colorimeter
 Instrument: Photometer
 Instrument: Spectrophotometer
 Integrating Sphere
 Luminance Meter
 Monochromator
 Photodetector
 Reflectance Standards

Spectral Power Distribution
 Spectrometer
 Standard Lamp
 Standard Measurement Geometries

People**Section Editor: Prof. Renzo Shamey**

Al-Biruni
 Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan
 Al-Haytham (Alhazen), Abū ‘Alī al-Ḥasan ibn
 al-Ḥasan ibn
 Al-Tusi, Nasir al-Din
 Avicenna, Abū ‘Alī al-Ḥusayn ibn ‘Abd Allāh ibn
 Sīnā
 Bartleson, C. James
 Billmeyer, Fred Wallace, Jr.
 Crawford, Brian Hewson
 Dalton, John
 Fechner, Gustav Theodor
 Goethe, Johann Wolfgang von
 Guild, John
 Helmholtz, Hermann Ludwig von
 Hering, Karl Ewald Konstantin
 Ibn Rushd (Averroes)
 Ibn Sahl, Abu Sa’d al-‘Ala’
 Itten, Johannes
 Ives, Frederic Eugene
 Judd, Deane Brewster
 Katz, David
 Kohlrausch, Arnt
 Kries, Johannes Adolph von
 Kubelka, Paul
 Lambert, Johann Heinrich
 Le Blon, Jacob Christoph
 Lippmann, Jonas Ferdinand Gabriel
 Lovibond, Joseph Williams
 MacAdam, David L
 Maxwell, James Clerk
 Mayer, Tobias
 Munsell, Albert Henry
 Newton, (Sir) Isaac
 Nickerson, Dorothy
 Ostwald, Friedrich Wilhelm
 Palmer, George
 Purkyně, Jan Evangelista
 Richter, Manfred

Runge, Philipp Otto
 Rushton, William A. H.
 Saltzman, Max
 Schrödinger, Erwin
 Stanziola, Ralph A
 Stevens, Stanley Smith
 Stiles, Walter Stanley
 Wright, David
 Wyszecki, Gunter

Physics

Section Editor: *Dr. Joanne Zwinkels*

Apparent Magnitude, Astronomy
 Blackbody and Blackbody Radiation
 Color Centers
 Color Magnitude Diagrams
 Compton Scattering
 Galaxy Color Magnitude Diagram
 Hertzsprung-Russell Diagram
 Iridescence (Goniochromism)
 Lanthanoid Ion Color
 Light, Electromagnetic Spectrum
 Luminescence
 Rayleigh and Mie Scattering
 Structural Color
 Surface Plasmons
 Transition Metal Ion Colors
 Umov Effect

Standards: CIE

Section Editor: *Prof. Ming Ronnier Luo*

CIE 1931 and 1964 Standard Colorimetric
 Observers: History, Data, and Recent
 Assessments
 CIE Chromatic Adaptation: Comparison of von
 Kries, CIELAB, CMCCAT97, and CAT02
 CIE Chromaticity Coordinates (xyY)
 CIE Chromaticity Diagrams, CIE Purity, CIE
 Dominant Wavelength
 CIE Color Rendering Index
 CIE Guidelines for Evaluation of Gamut Mapping
 Algorithms: Summary and Related Work
 (Pub. 156)

CIE Guidelines for Mixed Mode Illumination:
 Summary and Related Work
 CIE Method of Assessing Daylight Simulators
 CIE Physiologically Based Color Matching
 Functions and Chromaticity Diagrams
 CIE Special Metamerism Index: Change in
 Observer
 CIE Standard Illuminants and Sources
 CIE Tristimulus Values
 CIE u' , v' Uniform Chromaticity Scale Diagram
 and CIELUV Color Space
 CIE Whiteness
 CIE94, History, Use, and Performance
 CIECAM02
 CIEDE2000, History, Use, and Performance
 CIELAB
 CIELAB for Color Image Encoding (CIELAB,
 8-Bit; Domain and Range, Uses)
 Daylight Illuminants
 Spectral Luminous Efficiency
 USC Diagrams; Uniform Chromaticity Scales;
 Y_u/v'

Vision: Concepts

Section Editor: *Prof. Dr. Sophie Wuerger, Ph.D.*

Binocular Color Matching
 Color and Visual Search, Color Singletons
 Color Processing, Cortical
 Color-Opponency, Unique Hues
 Cone Fundamentals
 Pseudoisochromatic Plates
 Tetrachromatic Vision
 Visual Evoked Potentials

Vision: Concepts-2

Section Editor: *Prof. Yoko Mizokami*

Anchoring Theory of Lightness
 Chromostereopsis
 Color Contrast Adaptation
 Color Information Processing in Higher Brain
 Areas
 Color Scene Statistics, Chromatic Scene Statistics
 Environmental Influences on Color Vision

Helmholtz–Kohlrausch Effect
Hyperspectral Imaging
Melanopsin Contributions to Human Brightness Perception
Minimally Distinct Border
Motion and Color Cognition
Parafovea
Perceptual Grouping and Color
Retinex Theory
Scintillating Scotoma

Vision: Phenomena

Section Editor: *Prof. Rolf G. Kuehni*

Adaptation
Afterimage
Assimilation
Color Constancy
Color Perception and Environmentally Based Impairments
Color Phenomenology
Color Spreading, Neon Color Spreading, and Watercolor Illusion

Color Synesthesia
Color Vision Testing
Colors: Where Are They?
Deuteranopia
Fechner’s Colors and Benham’s Top
Ganzfeld
Mach Bands
Protanopia
Simultaneous Color Contrast
Transparency
Tritanopia

Vision: Physiology

Section Editor: *Prof. Rolf G. Kuehni*

Blind Spot
Evolution of Primate Color Vision
Ganglion Cells
Magno-, Parvo-, and Koniocellular Pathways
Melanopsin Retinal Ganglion Cells
Optic Chiasm, Chiasmal Syndrome
Photoreceptors, Color Vision

About the Editor



Renzo Shamey has been actively involved in the field of color science and technology since the early 1990s. His significant contributions in this domain encompass the development of datasets for color difference modeling, exploration of unique hues, blackness, whiteness, and investigation of sources of variability in perceptual assessments of color and color differences. Additionally, his research has focused on the automation of the coloration process and the implementation of artificial intelligence for diagnosing faults. He has also created undergraduate and graduate courses in color science and coloration, delivering lectures at prestigious universities in the United States, Europe, and Asia.

Professor Shamey has authored or co-authored over 250 peer-reviewed manuscripts, book chapters, monographs, and notable books such as *Modelling, Simulation, and Control of the Dyeing Process* (2014) and *Pioneers of Color Science* (2020). Furthermore, he has successfully supervised more than 30 PhD students who have made significant contributions to color science and technology.

Professor Shamey has actively participated in various color organizations including, but not limited to, the Society of Dyers and Colourists (SDC), the Inter-Society Color Council (ISCC), the American Association of Textile Chemists and Colorists (AATCC), and the International Commission on Illumination (CIE). Notably, he served as the president-elect, president, and past president of the ISCC from 2018 to 2022. He has also held positions such as the Chair of the Color Measurement Test Methods Committee (RA36)

at AATCC from 2011 to 2014 and has been a member of several technical committees.

In 2014, Dr. Shamey was selected as a North Carolina State University Faculty Scholar. His contributions have been acknowledged through several awards, including the Gold Medal from the Worshipful Company of Dyers for a research paper he co-authored (2009), the AATCC JW Weaver Award for the Best Paper of the Year (2012), and the Fellowship of the Society of Dyers and Colourists (2000). Currently, he is the CIBA distinguished Professor of Color Science and Technology at North Carolina State University in the United States and is also an Honorary Professor at Tiangong University in China.

Section Editors

Section: Art and Design



José Luis Caivano

Secretaria de Investigaciones FADU-UBA
Universidad de Buenos Aires, and Conicet
Buenos Aires, Argentina

Section: Materials: General



Harold S. Freeman

Wilson College of Textiles
North Carolina State University
Raleigh, NC, USA

Sections: Vision: Phenomena; Vision: Physiology



Rolf G. Kuehni

Color Science Laboratory
North Carolina State University
Raleigh, NC, USA

Section: Metrology: Concepts and Devices

Tzeng-Yow Lin Industrial Technology Research Institute, Hsinchu, Taiwan

Section: Standards: CIE

Ming Ronnier Luo

School of Design, University of Leeds
Leeds, UK

University of Zhejiang
Zhejiang, China

National Taiwan University of Science
and Technology
Taipei City, Taiwan

Section: Language and Categorisation

Lindsay MacDonald

Faculty of Engineering Sciences
University College London
London, UK

Section: Vision: Concepts-2

Yoko Mizokami

Graduate School of Engineering
Chiba University
Chiba, Japan

Section: Computer Graphics

Erik Reinhard
InterDigital
Cesson-Sévigné, France

Section: People

Renzo Shamey
Color Science and Imaging Laboratory
Wilson College of Textiles
North Carolina State University
Raleigh, NC, USA

Section: Color Engineering: Computations

Gaurav Sharma
University of Rochester
Rochester, NY, USA

Section: Lighting

Wout van Bommel
van Bommel Lighting Consultant
Nuenen, The Netherlands

Section: Color Science: Computations

Stephen Westland
Leeds Institute of Textiles and Colour
University of Leeds
Leeds, UK

Section: Vision: Concepts

Sophie Wuerger
Department of Psychology
University of Liverpool
Liverpool, UK

Section: Physics



Joanne Zwinkels

National Research Council Canada
Ottawa, ON, Canada

Contributors

Mira Abed Department of Textile Engineering, Chemistry & Science, North Carolina State University, Raleigh, NC, USA

Jonathan Aboshiha The UCL Institute of Ophthalmology and Moorfields Eye Hospital, London, UK

Arash Akbarinia FB06 Psychologie und Sportwissenschaften, Justus-Liebig Universität Giessen, Giessen, Germany

Nancy Alvarado Department of Psychology and Sociology, California State Polytechnic University, Pomona, CA, USA

Seyed Hossein Amirshahi AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

Malvina Arrarte-Grau Landscape and Color Design, Arquitectura Paisajismo Color, Lima, Peru

Landscape and Color Designer, Environmental Color Design Study Group (AIC), Lima, Peru

Rengin Aslanoğlu Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

Ulrich Bachmann Institute of Colour and Light, Zurich, Switzerland

Thomas Bechtold Research Institute for Textile Chemistry and Textile Physics, Leopold Franzens University of Innsbruck, Dornbirn, Vorarlberg, Austria

Tony Belpaeme School of Computing and Mathematics, University of Plymouth, Plymouth, UK

Nathan A. Benjamin Calit2, Computer Science, University of California, Irvine, Irvine, CA, USA

Martin Bide Department of Textiles, University of Rhode Island, Fashion Merchandising and Design, Kingston, RI, USA

David L. Bimler School of Psychology, Massey University, Palmerston North, New Zealand

Victoria Bogushevskaya Department of Humanities, University of Salento, Lecce, Italy

Valerie Bonnardel Psychology, University of Winchester, Winchester, UK

Alex Bourque Department of Chemistry, Dalhousie University, Halifax, NS, Canada

David Briggs Short Courses, National Art School, Sydney, NSW, Australia

Michael H. Brill Datacolor, Lawrenceville, NJ, USA

Berit Brogaard Department of Philosophy, University of Oslo, Oslo, Norway

Brogaard Lab for Multisensory Research, University of Miami, Miami, FL, USA

Angela M. Brown College of Optometry, Department of Optometry, Ohio State University, Columbus, OH, USA

Robert R. Buckley National Archives of the UAE, Abu Dhabi, UAE

John D. Bullough Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

José Luis Caivano Secretaría de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

Ellen C. Carter Color Research and Application and, Konica Minolta, Pennsville, NJ, USA

J. N. Chakraborty Department of Textile Technology, National Institute of Technology, Jalandhar, Punjab, India

Stephanie M. Chang Calit2, Computer Science, University of California, Irvine, Irvine, CA, USA

Mark Changizi 2ai Labs, Institute for the Study of Human and Machine Cognition, Columbus, OH, USA

Lisa Chapman North Carolina State University, Raleigh, NC, USA

Cheng-Hsien Chen Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Vladimir G. Chigrinov Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong
Department of Theoretical Physics, Moscow Region State University, Mytishi, Russia

Nanjing Jingcui Optical Technology Co., Ltd, Nanjing, China

Jae Hong Choi Department of Textile System Engineering, Kyungpook National University, Daegu, South Korea

Robert Christie School of Textiles & Design, Heriot-watt University Scottish Borders Campus, Galashiels, UK

Yi-Chen Chuang Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Filip Ciesielczyk Faculty of Chemical Technology, Institute of Chemical Technology and Engineering, Poznan University of Technology, Poznan, Poland

Richard S. Cook Department of Linguistics, University of California, Berkeley, CA, USA

Oswaldo Da Pos Department of General Psychology, University of Padua, Padua, Italy

Olga Daneyko Department of Psychology, Sociology and Politics, Sheffield Hallam University, Sheffield, UK

Jules Davidoff Department of Psychology, Goldsmiths University of London, London, UK

Don Dedrick Department of Philosophy and Department of Psychology, University of Guelph, Guelph, ON, Canada

Peter Dehoff Zumtobel Lighting, Dornbirn, Austria

Prutha S. Deshpande Department of Psychology, Ohio State University, Columbus, OH, USA

Cognitive Sciences, University of California, Irvine, CA, USA

Yi Ding Donghua University, Shanghai, China

Tuane Cristina dos Santos North Carolina State University, Raleigh, NC, USA

Igor Douven SND/CNRS/Sorbonne University, Paris, France

Oskar Elek Computer Graphics Department, Max-Planck-Institut Informatik, Saarbrücken, Germany

Ahmed El-Shafei Department of Textile Engineering, Chemistry & Science, North Carolina State University, Raleigh, NC, USA

Mark D. Fairchild College of Science, Rochester Institute of Technology, Rochester, NY, USA

Susan P. Farnand Program of Color Science, Rochester Institute of Technology, Rochester, NY, USA

Mariana Figueiro Lighting Research Center, Troy, NY, USA

Anna Franklin Psychology, University of Surrey, Surrey, UK

Harold S. Freeman Fiber and Polymer Science Program, North Carolina State University, Raleigh, NC, USA

Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Sergio Gago Calit2, School of Engineering, University of California, Irvine, Irvine, CA, USA

Elena Garces Multimodal Simulation Lab (MSLab), Universidad Rey Juan Carlos, Madrid, Spain

Dimitria Electra Gatzia Department of Philosophy, University of Akron, Akron, OH, USA

Karl Gegenfurtner Department of Psychology, Justus-Liebig University, Giessen, Germany
Department of Psychology, Giessen University, Giessen, Germany

Abhijeet Ghosh Department of Computing, Imperial College London, London, UK

Alan Gilchrist Department of Psychology, Rutgers University, Newark, NJ, USA

Edward J. Giorgianni Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

Adam Glaz Department of Modern Languages, Maria Curie-Skłodowska University (UMCS), Lublin, Poland

Simone Gori Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy
Developmental Neuropsychology Unit, Scientific Institute “E. Medea”, Lecco, Italy

Parikshit Goswami School of Design, University of Leeds, Leeds, West Yorkshire, UK

Paul Green-Armytage The School of Media, Creative Arts and Social Enquiry, Curtin University, Perth, WA, Australia

Thomas L. Griffiths Departments of Psychology and Computer Science, Princeton University, Princeton, NJ, USA

Ruiqian Guo Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Institute for Electric Light Sources, Fudan University, Shanghai, China

Geoffrey Hallas Department of Color Chemistry, University of Leeds, Leeds, UK

J. Richard Hanley Department of Psychology, University of Essex, Colchester, UK

C. L. Hardin Department of Philosophy, Syracuse University, Syracuse, NY, USA

Amanda Hardman School of Psychology, University of Aberdeen, Aberdeen, UK

Ian G. Harris Computer Science, University of California, Irvine, CA, USA

Mathieu Hébert ERIS Group, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet de Saint-Etienne, Saint-Etienne, France

Roger D. Hersch Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Robert Hirschler Colour Consultant, Budapest, Hungary

Donald Hoffman Department of Cognitive Sciences, University of California at Irvine, Irvine, CA, USA

Stephanie Huette Department of Psychology, University of Memphis, Memphis, TN, USA

Shao-Tang Hung Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Robert W. G. Hunt Department of Colour Science, University of Leeds, Leeds, UK

Kimberly A. Jameson Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA

Adrian Jarabo Graphics and Imaging Lab, Universidad de Zaragoza, I3A, Zaragoza, Spain

Teofil Jesionowski Faculty of Chemical Technology, Institute of Chemical Technology and Engineering, Poznan University of Technology, Poznan, Poland

Yang Jiao Computer Science, University of California, Irvine, CA, USA

Gabriele Jordan Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK

Yasmina Jraissati American University of Beirut, Beirut, Lebanon
Ronin Institute, NJ, USA

Paul Kay Department of Linguistics, University of California, Berkeley, CA, USA

Algy Kazlaucius Department of Color Science, University of Leeds, Leeds, West Yorkshire, UK

T. Q. Khanh Laboratory of Lighting Technology, Technische Universität Darmstadt, Darmstadt, Germany

Sung-Hoon Kim Department of Textile System Engineering, Kyungpook National University, Daegu, Republic of Korea

Daniel C. Kiper Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Eric Kirchner Kirchner Publications, AJ, Leiden, The Netherlands
Color Technology Department, AkzoNobel Paints and Coatings, Zuid-Holland, The Netherlands

Akiyoshi Kitaoka Department of Psychology, Ritsumeikan University, Kyoto, Japan

Martine Knoop Berlin University of Technology, Berlin, Germany

J. Kobbert Laboratory of Lighting Technology, Technische Universität Darmstadt, Darmstadt, Germany

Hidehiko Komatsu Brain Science Institute, Tamagawa University, Machida, Japan

Balázs Kránicz University of Pannonia, Veszprém, Hungary

John Krantz Natural Sciences, Department of Psychology, Hanover College, Hanover, IN, USA

Jan Kremers Department of Ophthalmology, University Hospital Erlangen, Erlangen, Germany

Rolf G. Kuehni Color Research & Application, Pennsville, NJ, USA
Charlotte, NC, USA

Color Science Laboratory, Carolina State University, Raleigh, NC, USA

Traci A. M. Lamar Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Eun-Mi Lee Department of Textile System Engineering, Kyungpook National University, Daegu, Republic of Korea

Kassandra R. Lee Department of Psychology, University of Nevada, Reno, NV, USA

David J. Lockwood Metrology Research Centre, National Research Council Canada, Ottawa, ON, Canada

Jorge Lopez-Moreno GMRV Group, Universidad Rey Juan Carlos, Madrid, Spain

Robert Lucas Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK

Ming Ronnier Luo School of Design, University of Leeds, Leeds, UK
State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

Lindsay W. MacDonald Faculty of Engineering Sciences, University College London, London, UK

Colin M. MacLeod Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Kevin Mansfield University College of London Institute for Environmental Design and Engineering, London, UK

John Mardaljevic School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, Leicestershire, UK

Jasna Martinovic Department of Psychology, University of Edinburgh, Edinburgh, UK

School of Psychology, University of Aberdeen, Aberdeen, UK

Jennifer Matey Department of Philosophy, Southern Methodist University, Dallas, TX, USA

Barry Maund Department of Philosophy, The University of Western Australia, Crawley, WA, Australia

John McCann McCann Imaging, Arlington, MA, USA

Terry McGowan Lighting Ideas, Inc, Cleveland Heights, OH, USA

Manuel Melgosa Optics Department, University of Granada, Granada, Spain

Ralf Michel Institute Integrative Design | Master Studio Design, University of Applied Sciences FHNW Basel, Basel, Switzerland

Galen Minah Department of Architecture, College of Built Environments, University of Washington, Seattle, WA, USA

Yoko Mizokami Graduate School of Engineering, Chiba University, Chiba, Japan

John D. Mollon Department of Experimental Psychology, Cambridge University, Cambridge, UK

Guillermo Monros Departamento de Química Inorgánica y Orgánica, Universitat Jaume I, Castelló, Spain

Nathan Moroney Hewlett-Packard Laboratories, Palo Alto, CA, USA

Jan Morovic Hewlett-Packard Company, Sant Cugat del Valles/Barcelona, Catalonia, Spain

Maria Luisa Musso Architecture, Design and Urbanism, Universidad de Buenos Aires, Color Research Program, Buenos Aires, Argentina

Dimitris Mylonas Department of Computer Science, University College London, London, UK

Antal Nemcsics Budapest, Hungary

Juan Luis Nieves Optics Department, Science Faculty, University of Granada, Granada, Spain

Larissa Noury Colour-Space-Culture, Paris, France

Leonhard Oberascher FH Joanneum University of Applied Sciences, Salzburg, Austria

Zena O'Connor Design Research Associates, University of Sydney, Sydney, NSW, Australia

Nilgün Olguntürk Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

Maria Olkkonen Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Department of Psychology, Durham University, Durham, UK

Gertrud Olsson Academy of Design and Crafts at Steneby, University of Gothenburg, Gothenburg, Sweden

Roy Osborne Color Research and Application, The Colour Group of Great Britain, London, UK

Gilles Page Signify, Miribel, France

Stephen E. Palmer Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Galina V. Paramei Department of Psychology, Liverpool Hope University, Liverpool, UK

C. Alejandro Parraga Computer Vision Centre/Computer Science Department, Universitat Autònoma de Barcelona, Bellaterra, Spain

Neil Parry Vision Science Centre, Manchester Royal Eye Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Centre for Ophthalmology and Vision Sciences, Institute of Human Development, University of Manchester, Manchester, UK

Emily J. Patterson Institute of Ophthalmology, University College London, London, UK

Baingio Pinna Department of Biomedical Sciences, University of Sassari, Sassari, Italy

Renata Pompas Milan, Italy

Bao-Jen Pong Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Tania Pouli Advanced Media Contents Lab, IRT b-com, Cesson-Sévigné, France

Maria Manuela Marques Raposo Department of Chemistry, University of Minho, Braga, Portugal

Peter Raynham University College of London Institute for Environmental Design and Engineering, London, UK

Erik Reinhard InterDigital, Cesson-Sévigné, France

Tobias Ritschel Department of Computer Graphics, MPI Informatik, Saarbrücken, Germany

Herzog Robert Max-Planck Institute for Informatics, Saarbrücken, Germany

Marisa Rodriguez-Carmona Optometry and Visual Science, City University, London, London, UK

Georges Roque Centre National de la Recherche Scientifique (CNRS), Paris, France

Ken Sagawa Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Masato Sakurai Shizuoka Institute of Science and Technology, Fukuroi, Japan

Abhijit Sarkar Mixed Reality Systems Engineering, Microsoft Corporation, Redmond, WA, USA

P. R. B. Saull Metrology Research Centre – Ionizing Radiation Standards, National Research Council, Ottawa, ON, Canada

János Schanda Veszprém, Hungary

Karen B. Schloss Department of Psychology and Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, WI, USA

Eriko Self Department of Psychology, California State University, Fullerton, CA, USA

Renzo Shamey Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

David R. Simmons School of Psychology, University of Glasgow, Glasgow, Scotland, UK

A. Soranzo Faculty of Development and Society, The Department of Psychology, Sociology and Politics, Sheffield Hallam University, Sheffield, UK

Andrew Stockman Department of Visual Neuroscience, UCL Institute of Ophthalmology, London, UK

Kartic Subr Department of Computer Science, University College London, London, UK

Suchitra Sueeprasan Department of Imaging and Printing Technology, Chulalongkorn University Intellectual Repository, Chulalongkorn University, Bangkok, Thailand

Lois Swirnoff The Cooper Union for the Advancement of Science and Art, New York, NY, USA

Malgorzata Szymczyk Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Yoshika Takahashi Graduate School of Engineering, Chiba University, Chiba, Japan

Ken Tapping D.R.A.O, National Research Council Canada, Penticton, BC, Canada

Muneto Tatsumoto Medical Safety Management Center, Dokkyo Medical University Hospital, Tochigi, Japan

Sean Tauber Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA

Li-Lin Tay Measurement Science and Standards, National Research Council Canada, Ottawa, ON, Canada

Krzysztof Templin Computer Graphics, Max Planck Institute for Informatics, Saarbrücken, Germany

Peter Thorns Strategic Lighting Applications, Thorn Lighting Ltd, Durham, UK

Richard J. D. Tilley Queen's Buildings, Cardiff University, Cardiff, UK

Andrew Towns Lambson Ltd., Wetherby, West Yorkshire, UK

Katherine E. M. Tregillus Center for Cognitive Sciences, University of Minnesota, Minneapolis, USA

H. Joel Trussell Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

Sei-ichi Tsujimura Graduate School of Design and Architecture, Nagoya City University, Aichi, Japan

Begüm Ulusoy Interior Architecture and Design, School of Design, University of Lincoln, Lincoln, Brayford Pool, UK

Gisela Umbuzeiro School of Technology, UNICAMP, Limeira, SP, Brazil
Fiber and Polymer Science Program, North Carolina State University, Raleigh, NC, USA

Philipp Urban Fraunhofer Institute for Computer Graphics Research IGD, Technische Universität Darmstadt, Darmstadt, Germany

Mari Uusküla School of Humanities, Tallinn University, Tallinn, Estonia

Arne Valberg Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

Wout van Bommel van Bommel Lighting Consultant, Nuenen, The Netherlands

Peter van der Burgt Philips Lighting, Valkenswaard, The Netherlands

Michal Vik Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic

Poorvi L. Vora Department of Computer Science, The George Washington University, Washington, DC, USA

Mark Wainwright Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK

David Alan Warburton Institute for the History of Ancient Civilizations, Northeast Normal University, Changchun, China

Romann M. Weber DisneyResearch|Studios, Zurich, Switzerland

Michael A. Webster Department of Psychology, University of Nevada, Reno, NV, USA

Chang Wei Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Institute for Electric Light Sources, Fudan University, Shanghai, China

Bor-Jiunn Wen Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Chao-Hua Wen Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Stephen Westland Colour Science and Technology, University of Leeds, Leeds, UK

Mary Anne White Department of Chemistry, Dalhousie University, Halifax, NS, Canada

Tova N. Williams Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Jonathan Winawer Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

Nathan Witthoft Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

Department of Psychology, Stanford University, Stanford, CA, USA

Christoph Witzel School of Psychology, University of Southampton, Southampton, UK

Wayne Wright Department of Philosophy, California State University, Long Beach, CA, USA

Sophie Wuerger Department of Psychology, University of Liverpool, Liverpool, UK

Bei Xiao Department of Computer Science, American University, Washington, DC, USA

Fengxian Xie Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Institute for Electric Light Sources, Fudan University, Shanghai, China

Hirohisa Yaguchi Emeritus, Chiba University, Chiba, Japan

Noga Zaslavsky Department of Brain and Cognitive Sciences and Center for Brains Minds and Machines, MIT, Cambridge, MA, USA

Daniele Zavagno Department of Psychology, University of Milano-Bicocca, Milan, Italy

Suncica Zdravkovic Department of Psychology, University of Novi Sad, Novi Sad, Serbia

Wanlu Zhang Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Institute for Electric Light Sources, Fudan University, Shanghai, China

Jiyu Zhong National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Weidong Zhuang National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Joanne Zwinkels National Research Council Canada, Ottawa, ON, Canada

A

Achromatic Color

- [Anchoring Theory of Lightness](#)

Achromatopsia

- [Motion and Color Cognition](#)

Acquired Color Vision Impairment

- [Color Perception and Environmentally Based Impairments](#)

Acquired Color Vision Loss

- [Color Perception and Environmentally Based Impairments](#)

Adaptation

Susan P. Farnand

Program of Color Science, Rochester Institute of Technology, Rochester, NY, USA

Definition

The human visual system receives and interprets an enormous volume of information regarding the environment in which the individual is immersed. Indeed, there is more information available than can be processed. This includes information on whether it is day or night, sunny or cloudy, how fast objects around us are moving and whether an approaching person is happy or angry. A degree of normalization of human vision is of critical importance in establishing in the brain and mind a tool for recognizing objects or people despite the extensive variations in physical stimuli that the visual system encounters. The term used to identify this tool and its many aspects is “adaptation.” Adaptation processes occur throughout the visual system, from the pupil to the brain, with the results accessible to the conscious mind. Webster cites the key functions of adaptation as efficient coding of the incoming visual information, including enhancement of novel stimuli, and appearance needed for

recognition [1]. The general subject of adaptation also applies to sensory systems other than vision.

Adaptation

Visual adaptation processes have developed over many millions of years by evolutionary processes, with the details varying between different advanced species of creatures, depending on their visual systems and the world they inhabit. The human adaptation system is known to be complex in detail with many aspects not yet fully understood. One of its purposes is to adapt incoming stimuli representing known objects in a way that the perceptual appearance of these remains very similar for most natural conditions in which they are viewed. Adaptation is understood as an efficient way to code physical information entering the visual system through the eyes in a manner that results in a reasonably constant perceived appearance of known people and objects while retaining enough information to preserve the context in which we find them.

Adaptation processes are applied by the visual system at different stages of the viewing process, thereby having potential impact on all perceptual aspects, from the retina in the eyes to the mind via various levels of activity in the brain, influencing the final perceptions in the mind. Important functions of adaptation at earlier stages are enhancing discrimination, highlighting visual novelties, maintaining perceptual constancy, producing efficient vision coding, getting extensive information from neurons with very limited dynamic ranges, and ensuring that neurons at lower levels in the system are not under- or overexposed [2]. Fairchild [3] lists a number of mechanisms from the eye to the mind that are thought to be involved in the adaptation process including pupil diameter, depletion and regeneration of photopigments, transition from cones to rods, pooling of photoreceptor input, opponency effects, gain control in receptors and higher-order mechanisms, and cognitive effects.

In the International Lighting Vocabulary of the *Commission Internationale de l'Eclairage* (CIE), the term adaptation is defined as "Process by which the state of the visual system is modified

by previous and present exposure to stimuli that may have various luminance values, spectral distributions, and angular subtenses" [4]. This definition suggests that modifications to the visual system can occur due to shifts in luminance (dark and light adaptation) and spectral distribution (chromatic adaptation). These visual adaptation processes are significant for human color perception and are briefly discussed below.

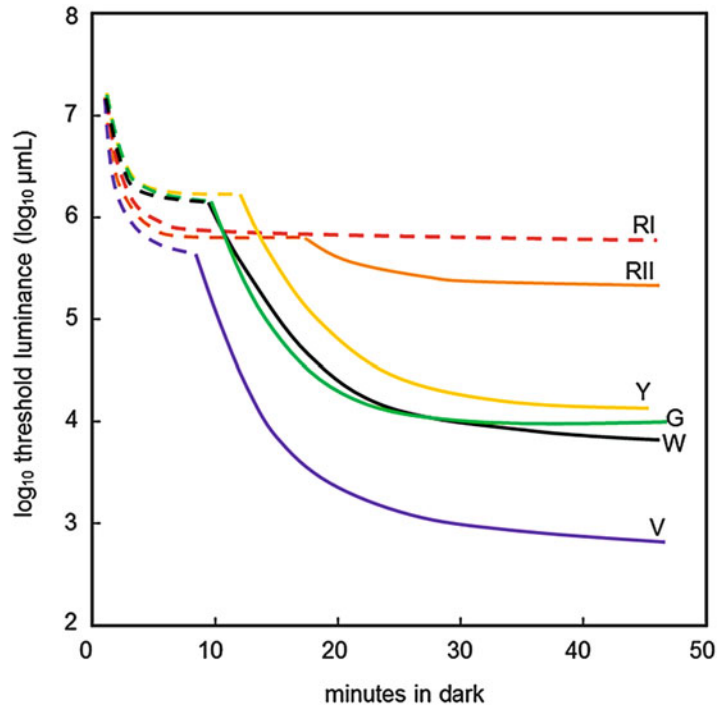
A thorough, recent review by Webster of the present knowledge about visual adaptation, as well as key current questions pertaining to it, is found in Ref. [1].

Dark and Light Adaptation

Both concepts deal with the fact that natural light levels on the earth can vary more than a million-fold in intensity during a 24-h day cycle (from a dark night to a bright, sunny day). Our visual system can compensate for small changes in light level very quickly so that we are not aware of them; when the changes are large, such as when moving from a dark, windowless room into bright daylight (or vice versa), the visual system may take up to an hour to completely adjust to the change. The range of response of neurons to light stimuli, for example, in the ganglion cells of the retina, runs from zero to about 200 spikes per second. This is a very small range to account for the required millionfold range in light intensity. Some control systems have developed to overcome this limitation: one is the pupil size in the eyes that can range in diameter from about 2 to 8 mm; it increases the area in which light can enter an eye by a modest factor of about 16. Another supporting system is having two types of sensors in the eye with much different sensitivities to light. These are the rods, active in night vision, and the cones, active in daylight vision. Depending on the light intensity, one or both can operate, both being active in twilight illumination.

Dark adaptation refers to the situation when an observer transitions from exposure to bright light to darkness. If this transition is abrupt, the ability to perceive objects and their differences in color disappears initially and nearly completely. Figure 1 illustrates how the visual sensitivity

Adaptation, Fig. 1 Dark adaptation curves for test stimuli of various wavelengths: RI (680 nm), RII (635 nm), Y (573 nm), G (520 nm), V (485 nm), and W (white light). (Chapanis' data from Bartlett [6])



threshold changes during dark adaptation following a photoreceptor bleaching exposure. The upper part of the curves represents the recovery of the cone receptors, showing a rapid drop in threshold over several minutes before leveling off. After about 10–15 min, depending on the spectral content of the stimulus, the rod photoreceptors recover enough sensitivity (except for the longest wavelength stimulus) such that visual system shifts to rod receptor operation, lower part of the curves. For a white light stimulus, the break point can vary between about 7 and 10 min, depending on the observer [5]. Over 20 min of recovery time are required to obtain optimal visual sensitivity of the rod receptor vision system.

Light adaptation describes the opposing situation where the observer moves from exposure to darkness into brightly lit surroundings with the resulting change in visual sensitivity again measured as it changes over time in the bright surroundings. In these conditions the time for recovery of visual sensitivity is much shorter (6–8 min) than in dark adaptation because the cone sensory system recovers more quickly [7].

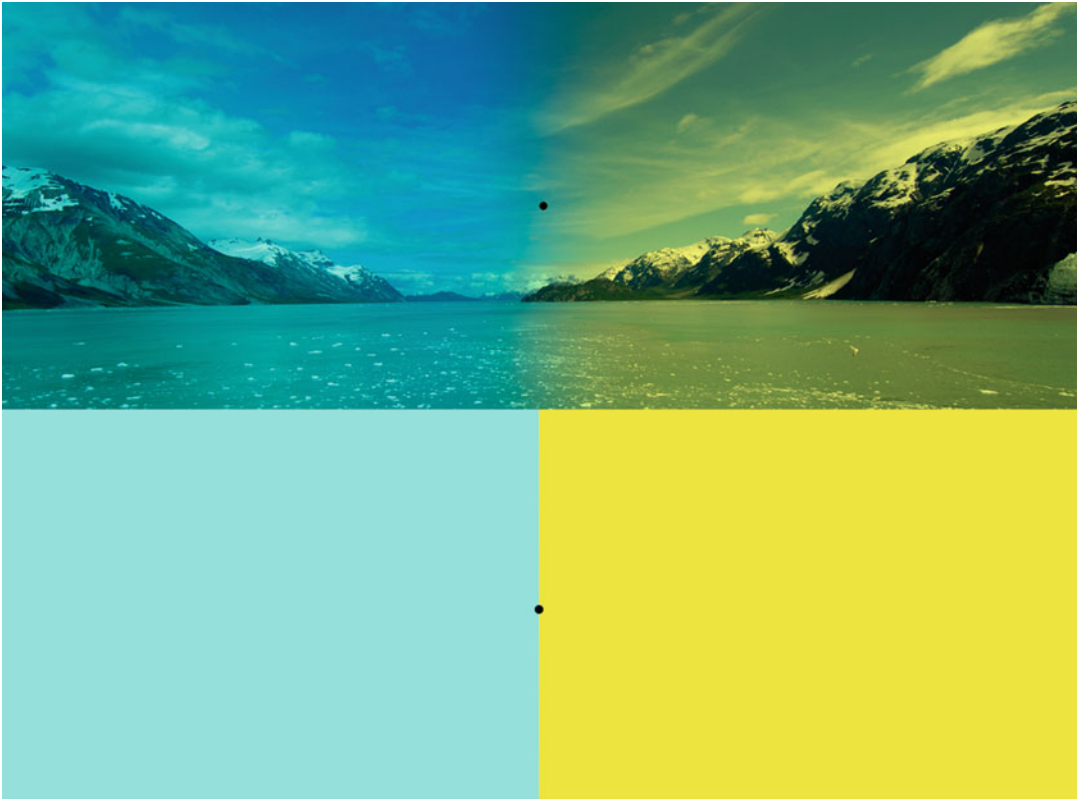
There are a number of factors influencing dark and light adaptation with the results, in terms of

visual sensitivity, varying noticeably. These can include environmental factors, such as overall light level and saturation level of the adapting stimuli, and characteristics of the observer, particularly age. Based on the large number of variables that can have an effect on adaptation, some researchers have created “adaptation metamers,” different stimuli that produce the same level of dark or light adaptation, in order to more precisely study the adaptation processes [1].

Given the considerable effect that variation in luminance or lightness has on the appearance of colored lights or objects, mathematical color appearance models often include adjustments to produce appearance predictions comparable to the references stored in the observer’s brain [2].

Chromatic Adaptation

Interaction of light with the sensor photopigments results in their bleaching to a smaller or larger degree, requiring a process of regeneration that takes some time. Humans have three kinds of cones in their visual systems, which allows relative adjustment in color perception in different



Adaptation, Fig. 2 An illustration of spatially localized chromatic adaptation. Stare at the dot between the blue and yellow patches for at least 15 s without moving your eyes. Now look at the white area. What do you see? Adapting to yellow fatigues the photoreceptors relative to their response to yellow, resulting in a bluish appearance. The

environments. Such relative temporary bleaching of the three cone types is also responsible for the dramatic afterimage effects seen when chromatic color samples are fixated for some period of time with the fixation then changing to a well-lit white

opposite occurs for blue. Repeat the process, this time, after adapting to the patches, look at the black dot in the photograph. By adjusting the adaptation time, you should be able to make the two halves of the photograph blend together

paper. Fields perceived originally as having a blue color now appear as having a yellow color and with comparably opponent color pairs in other cases. These afterimages also illustrate that chromatic adaptation is spatially localized on the

retina, sometimes referred to as local adaptation. In normal viewing conditions, the bleaching effects are much smaller, and in color perception these effects are normalized by an adaptation process.

That humans have three kinds of cone photoreceptors makes it possible for them to distinguish between a huge number of varying light compositions arriving at the eyes, and thereby also between objects. This situation requires chromatic adaptation so that the perception of known people and objects is adjusted to be similar to what is stored in the mind/brain and they are not considered new and different objects or people. In addition to relative bleaching of the three cone types, this adaptation can be achieved through relative gain control in the horizontal, bipolar, and ganglion cells in the retina as well as higher-level mechanisms beyond the retina [3].

An example of chromatic adaptation, as described by Fairchild [3], is that a piece of white-appearing paper viewed in average daylight still appears white when viewed in tungsten light, even though the colorimetric definition of the reflected energy absorbed by the eyes has changed considerably, as would also be the case if the paper is viewed in various other illumination colors. These results are based on chromatic adaptation effects automatically active in the normal human brain.

Over more than a century, a number of mathematical models have been developed to predict chromatic adaptation situations on basis of colorimetric data. One of the first and most influential models was introduced by J. von Kries (1902). Since then several new models have been proposed, among them the Land Retinex theory model (1977), the Nayatani model (1981), Guth's model (1985), and Fairchild's model (1990). All of these are based on L, M, S cone response data. Starting in the late 1990s, the CIE began to propose complex color appearance models that include chromatic adaptation, beginning with CIECAM97. An improvement of the latter was issued as CIECAM02 in 2002 [8]. This model is still in use though it has been further optimized. Detailed descriptions of these and additional models are found in Ref. [3].

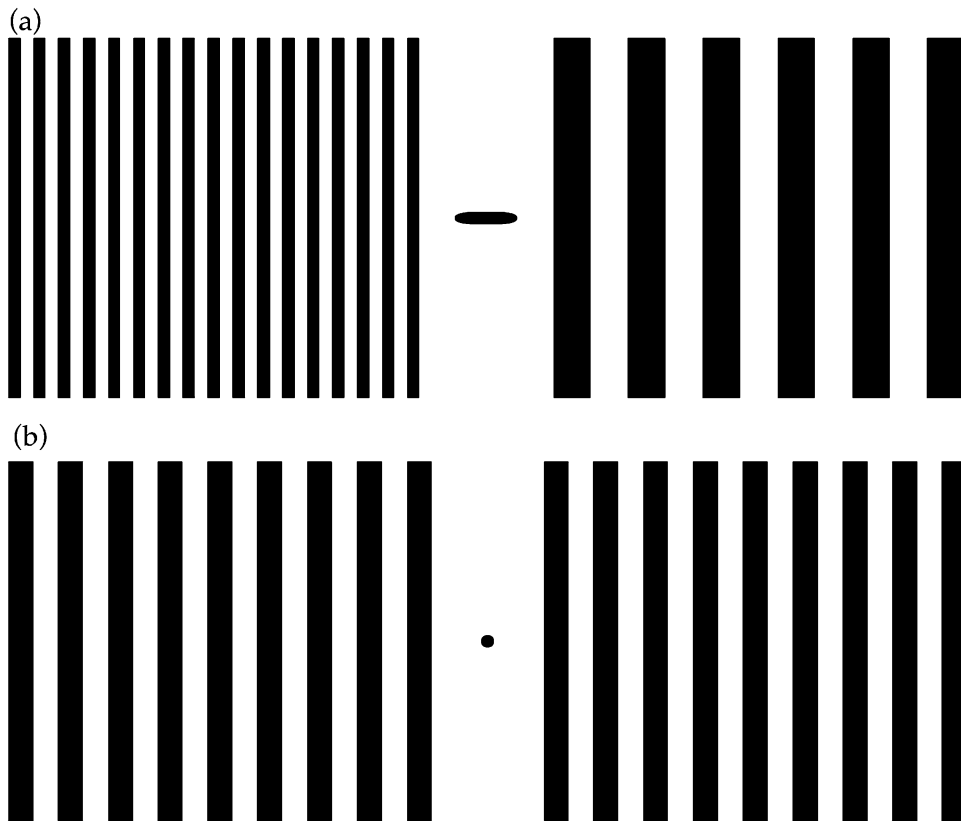
Spatial Adaptation

In addition to colorimetrically related adaptation, humans adapt to spatial characteristics including motion, blur, optical noise, and facial expression. An example of spatial adaptation is shown in Fig. 3. These have been demonstrated to exist in the human visual system, thereby assuring that related information is adjusted in the brain and mind, such as eliminating unrecognizable information due to motion, to a blurred image, or to general optical noise and, in facial images, assuring that the person involved is recognized correctly. But the neural processes in the human visual system that underlie adaptation are not yet fully identified and understood. This is on the list of future issues Webster includes in his review of visual adaptation [1]. The list also includes questions with regard to the limits of adaptation and an emphasis on evaluating how observers adapt to a more realistic visual world, as opposed to laboratory environments, including more complex visual stimuli. Snow et al. provide a review of research being conducted on visual adaptation with natural stimuli [9]. And McIntosh et al. present research on using convolutional neural networks to model retinal responses to natural scenes [10].

One additional topic of research pertains to the possibility of individual differences in adaptation. With the known differences in color matching functions along with the cognitive effects involved in adaptation, it seems reasonable to expect that individual observers would experience differences in how they adapt to given environments. Indeed, a computational study suggests the possibility for statistically significant differences in the adaptation processes among observers [11]. However, more extensive experimental data on metameric color matching and corresponding colors is needed to evaluate the extent of individual differences in adaptation.

Summary

Adaptation is an essential tool for the efficient and effective processing of the information available to human sensory systems. As Webster summarized the issue, "Adaptation operates throughout



Adaptation, Fig. 3 This figure provides an illustration of spatial adaptation. Direct your gaze to the horizontal black bar between the two sets of vertical bars in (a). Move your eyes back and forth along the bar for about a minute. Now fixate on the dot between the two sets of vertical bars in (b).

What do you see? The bars on the left of (b) should appear to be closer together than those on the right although the two sets of bars are identical. (After Fairchild Chapter 8, Figure 8.5 [3])

most, if not all, levels of sensory coding and thus impacts all aspects of perception. Adaptation probably serves a variety of related functions, ranging from increasing coding efficiency to supporting perceptual constancy and highlighting novel information” [1]. It is evident that more research is required to develop a more complete understanding of its processes (Fig. 2).

Cross-References

- [Afterimage](#)
- [CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02](#)
- [CIECAM02](#)
- [Color Appearance](#)
- [Color Vision Testing](#)

References

1. Webster, M.A.: Visual adaptation. *Ann. Rev. Vis. Sci.* **1**, 547–567 (2015)
2. Webster, M.A.: Probing the functions of contextual modulation by adapting images rather than observers. *Vis. Res.* **104**, 68–79 (2014)
3. Fairchild, M.D.: *Color Appearance Models*, 3rd edn. Wiley, Chichester (2013), Chapter 8
4. CIE: *International Lighting Vocabulary*, 2nd edn. Available at the online standard store <https://www.techsheet.com/cie> (2016)
5. Hecht, S., Mandelbaum, J.: The relation between vitamin A and dark adaptation. *J. Am. Med. Assoc.* **112**, 1910–1923 (1938)
6. Bartlett, J.N.R.: Chapter 8: dark and light adaptation. In: Graham, C.H. (ed.) *Vision and Visual Perception*. Wiley, New York (1965)
7. Kallionatis, M., Luu, C.: Light and dark adaptation. <https://webvision.med.utah.edu/book/part-viii-gabac-receptors/light-and-dark-adaptation/> (2019)

8. Fairchild, M.D.: Color Appearance Models, 3rd edn. Wiley, Chichester (2013)., Chapter 16
9. Snow, M., Coen-Cagli, R., Schwartz, O.: Adaptation in the visual cortex: a case for probing neuronal populations with natural stimuli. *F1000Research*. **6**, (2017)
10. McIntosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S., Baccus, S.: Deep learning models of the retinal response to natural scenes. *Adv. Neural Inf. Proces. Syst.* 1369–1377 (2016)
11. Fairchild, M. D.: Individual differences in color matching and adaptation: theory and practice. In: *Color and Imaging Conference*, vol. 2016, no. 1, pp. 115–120. Society for Imaging Science and Technology (2016)

Additivity Failure for Brightness

- [Helmholtz–Kohler Effect](#)

Advanced Front Lighting Systems (AFS) – Intelligent Light

- [Automotive Front Lighting System \(Status 2020\)](#)

Aftereffect

- [Afterimage](#)

Afterimage

Simone Gori
 Department of Human and Social Sciences,
 University of Bergamo, Bergamo, Italy
 Developmental Neuropsychology Unit, Scientific
 Institute “E. Medea”, Lecco, Italy

Synonyms

[Aftereffect](#); [Ghost image](#); [Image burn-in](#)

Definition

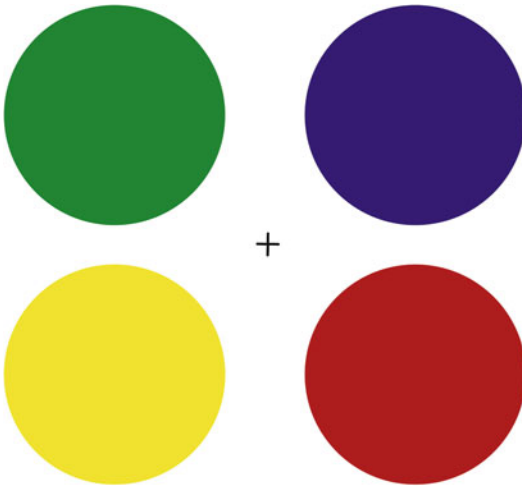
An afterimage is an image that continues to be perceived after the physical stimulus that it originated from disappears from the observer’s visual field. The afterimage can be a result of an exposition to a grayscale pattern, a colored pattern, or even to a motion stimulus. There are different kinds of afterimages, for example, the positive one will be the same with the physical stimulus, while the negative one will result in the opposite in terms of luminance, colors, or direction of the physical stimulus that generates it.

Introduction

In everyday experience it is not so uncommon to experience several types of aftereffects. The observation done by Addams [1] while watching a waterfall near to the Loch Ness in Scotland is probably the most famous description of the motion aftereffect outside a scientific laboratory. The motion aftereffect was already observed by Aristotle and Lucretius in the ancient times, but the description done by Addams remains the most cited example: if an observer watches a fall for about 60 s and, after that, moves the gaze to the woods nearby, everything will seem to move upward. Now it is known that motion aftereffects can be so powerful that they can influence the apparent direction of the illusory motion produced by static patterns, [2] even if generated after only milliseconds of adaptation [3].

Possible Explanations

However, it is not just adaptation to motion that produces aftereffects. Although an afterimage can be easily experienced after looking for a prolonged period of time at a grayscale image, the most popular afterimages are probably experienced after looking at a colored image. Staring at the fixation cross in Fig. 1 for more than 30 s and then moving the gaze to a white surface should produce a color negative afterimage: the disks will appear filled with the complementary colors of the



Afterimage, Fig. 1 In order to experience a color negative afterimage, the reader should stare at the fixation cross shown for no less than 30 s and then move the gaze to a white surface. What should be seen on the blank surface are four circles filled with the opposite colors (the *green* should be now *red*, the *blue* *yellow*, the *yellow* *blue*, and finally the *red* *green*)

original stimulus on the blank surface. The common opinion in the scientific community is that the afterimages result from neural adaptation, but adaptation is present throughout the visual system, which makes the neural substrate of these phenomena not so easy to isolate. For long, it was supposed that the afterimages were caused by fatigued cells in the retina responding to light. Then, if an observer will stare at the red color for enough time, the receptors in the retina responding to red will fatigue and will fire less. In color vision, it is generally believed that colors are represented along orthogonal opponent axes: yellow to blue and red to green. Consequently, when the observer will switch over to a white surface, after that the retinal receptors for red stimuli are adapted, the visual system will interpret that blank surface as filled with the red complementary color: green. This oversimplified explanation where all the color afterimages are explained at the retinal level was successively challenged by several phenomena that required cortical elaboration [4]. In the 1960s it was shown that the rabbit visual cortex presented

neural adaptation after prolonged exposition to a stimulus that was comparable to the duration of the aftereffects recorded during the psychophysical experiments [5]. These results, in agreement with other neuropsychological and psychophysical studies [4], moved the possible main physiological substrate of the afterimage into the primary visual cortex (V1). V1 is also able to explain the interocular transfer of the aftereffect, which is the possibility to adapt an eye and to observe some effects on the non-adapted one.

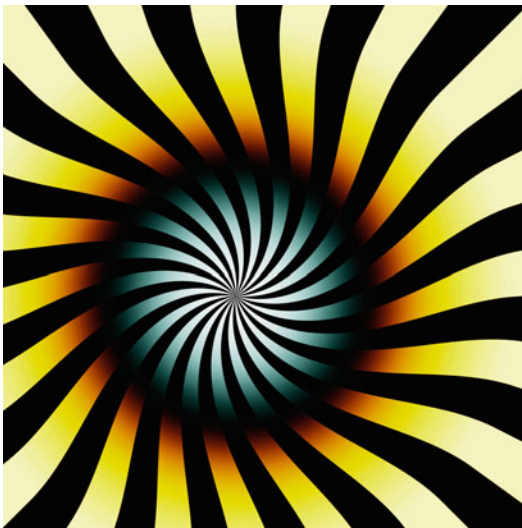
However, even if low-level cortical processes seem perfectly appropriate to explain the afterimage, the discovery that the top-down processes are able to influence radically both the primary visual cortex and also some subcortical “stations” of the visual system, such as the lateral geniculate nucleus (LGN), together with some new types of aftereffects that show the influence of attention [6, 7], consciousness [8], or an effect that lasts for days, weeks, or even months, seemed to clearly reject an explanation in terms of the short-lived adaptation reported typically in single cells of V1. These effects all suggested that the final percept of the afterimages is probably a complex result of the interaction and the adaptation of several subsystems throughout the visual system from the retina to the high-level cortical areas. Recently it was shown that the color afterimage signals are generated in the retina, but they are modified by cortical processes [9]. In sum the retinal afterimages face a destiny similar to that of all the retinal signals: in humans they are virtually always heavily modulated by the cortex.

An interesting question that can be posed, however, is whether the afterimages have a function or if they are just a side effect incurred as a minor consequence of having evolved such a complex and superb machinery known as the human visual system. The answer, as probably expected, seems to be that the afterimages, and neural adaptation in general, are part of complex neural strategies evolved during thousands of years in order to improve our perception of the world (e.g., [10, 11]).

Breathing Light Illusion

Finally the afterimage (with color or in grayscale, it depends on which version is used) is also at the basis of the family of illusions composed of the breathing light illusion and the dynamic luminance-gradient illusion [12–14]. Approaching these patterns like the one in Fig. 2 by moving one's head toward it makes the spot appear to become larger, more diffuse, and filled with white. On receding from it, the spot's center remains white, but the remainder appears smaller, darker, and sharper. The proposed explanation of the phenomenon is related to the superimposition of the afterimage on the physical stimulus during dynamical viewing [13, 14].

In summary the afterimages are interesting phenomena that help to understand some of our brain mechanisms. They are able to produce striking illusory effects. The underlying mechanisms appear to be relevant in producing a more efficient perception of the world.



Afterimage, Fig. 2 A complex version of the breathing light illusion. Approaching the patterns by moving one's head toward the image makes the spot appear to become larger, more diffuse, and filled with white. On receding from it, the spot's center remains white, but the remainder appears smaller, darker, and sharper. In addition, in this specific pattern, an illusory rotation is perceived together with the typical illusory effect of the breathing light illusion. This illusory rotation is the same as experienced in the Rotating-Tilted-Lines Illusion [15–18]

Cross-References

- ▶ [Adaptation](#)
- ▶ [Assimilation](#)
- ▶ [Simultaneous Color Contrast](#)

References

1. Addams, R.: An account of a peculiar optical phenomenon seen after having looked at a moving body etc. *London Edinburgh Philosoph. Magaz. J. Sci.* **5**, 373–374 (1834). 3rd series
2. Gori, S., Hamburger, K., Spillmann, L.: Reversal of apparent rotation in the Enigma-figure with and without motion adaptation and the effect of T-junctions. *Vis. Res.* **46**, 3267–3273 (2006)
3. Pavan, A., Skujevskis, M.: The role of stationary and dynamic test patterns in rapid forms of motion after-effect. *J. Vis.* **13**, 10 (2013)
4. Thompson, P., Burr, D.: Visual aftereffects. *Curr. Biol.* **19**, R11–R14 (2009)
5. Barlow, H.B., Hill, R.M.: Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature*. **200**, 1345–1347 (1963)
6. Reavis, E.A., Kohler, P.J., Caplovitz, G.P., Wheatley, T.P., Tse, P.U.: Effects of attention on visual experience during monocular rivalry. *Vis. Res.* **83**, 76–81 (2013)
7. Van Lier, R., Vergeer, M., Anstis, S.: Filling-in after-image colors between the lines. *Curr. Biol.* **19**, R323–R324 (2009)
8. van Boxtel, J.J.A., Tsuchiya, N., Koch, C.: Opposing effects of attention and consciousness on afterimages. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8883–8888 (2010)
9. Zaidi, Q., Ennis, R., Cao, D., Lee, B.: Neural locus of color afterimages. *Curr. Biol.* **22**(3), 220–224 (2012)
10. Greenlee, M.W., Heitger, F.: The functional role of contrast adaptation. *Vis. Res.* **28**, 791–797 (1988)
11. Barlow, H.B., Földiák, P.: Adaptation and decorrelation in the cortex. In: Darbin, R., Miall, C., Mitchison, G. (eds.) *The computing neuron*, pp. 54–72. Wesley Publishers, Reading (1989)
12. Gori, S., Stubbs, D.A.: A new set of illusions – the dynamic luminance gradient illusion and the breathing light illusion. *Perception*. **35**, 1573–1577 (2006)
13. Anstis, S., Gori, S., Wehrhahn, C.: Afterimages and the breathing light illusion. *Perception*. **36**, 791–794 (2007)
14. Gori, S., Giora, E., Agostini, T.: Measuring the breathing light illusion by means of induced simultaneous contrast. *Perception*. **39**, 5–12 (2010)
15. Gori, S., Hamburger, K.: A new motion illusion: the rotating-tilted-lines illusion. *Perception*. **35**, 853–857 (2006)
16. Gori, S., Yazdanbakhsh, A.: The riddle of the rotating tilted lines illusion. *Perception*. **37**, 631–635 (2008)

17. Yazdanbakhsh, A., Gori, S.: A new psychophysical estimation of the receptive field size. *Neurosci. Lett.* **438**, 246–251 (2008)
18. Gori, S., Mascheretti, S., Giora, E., Ronconi, L., Ruffino, M., Quadrelli, E., Facoetti, A., Marino, C.: The DCDC2 intron 2 deletion impairs illusory motion perception unveiling the selective role of magnocellular-dorsal stream in reading (dis)ability. *Cereb. Cortex.* **25**(6), 1685–1695 (2015)

Age-Related Changes of Human Seeing Abilities

► Vision and Age

Al-Biruni

Eric Kirchner
Kirchner Publications, AJ, Leiden,
The Netherlands

Abu'l-Raihan al-Biruni was a Persian polymath born in 973 AD. His contributions to mathematics, geography, astronomy, and physics are most well known. During his travels in India, he not only taught Greek science, but in exchange he became acquainted with Indian science. This made him a strong promoter of this science in the Islamic world.

At the court of sultan Mas'ud, al-Biruni wrote an important work on minerals and gemstones, called the *Kitab al-jamahir fi ma'rifat al-jawahir* (The book of the multitude of knowledge of precious stones) [1–3]. In this lapidary, al-Biruni gives a detailed description of the colors of many minerals and gemstones. Color is used as a clear way to identify minerals and gemstones. He discusses extensively the slight color differences between minerals originating from different mines or having different degrees of purity, relating them to the effect this may have on their financial value.

The *Kitab al-jamahir* does not contain a separate chapter on color itself or on color ordering. But taken together, the many descriptions of color

variations that appear in this book while discussing different types of minerals and gemstones do form a large body of scientific knowledge on color. Two centuries after al-Biruni, another Persian scientist by the name of al-Tusi would indeed collect many of al-Biruni's findings and formulate an impressive color ordering scheme by combining them with other sources [4].

An example of this part of al-Biruni's scientific legacy is found in his description of different types of rubies. According to al-Biruni, the colors of rubies range from sky blue, via lapis lazuli and indigo-blue, to kohl black. Almost exactly the same color series is found in the grand color scheme of al-Tusi, with only the color turquoise having been added. In a similar way, not only many color words in al-Tusi's grand color scheme already occur in al-Biruni's work, but the latter also provided several partial color orderings.

Another interesting aspect of al-Biruni's description of color is that he is one of the first to verbally describe one of the color dimensions. After having described the color series from sky blue, via lapis lazuli and indigo-blue, to kohl black, al-Biruni characterizes this gradual color change by the word *shab'a* (p. 72 in Ref. [1]) (Vol. 2, p. 42 in Ref. [5]). Originally this word is used for describing the feeling one has after a copious meal, and it is best translated as saturation. In classical Arabic literature and in modern Arabic, the same word became the common word for color saturation (p. 701 in Ref. [6]). Although the color sequence mentioned above shows that the colors in this series do not change only with respect to color saturation in the modern definition of this word, al-Biruni's text is one of the first to relate this term to color [7].

References

1. Krenkov, F.: *Al-jamahir fi ma'rifat al-jawahir*. Deccan, Hyderabad (1936)
2. al-Hadi, Y.: *Al-jamahir fi al-jawahir*. Scientific and Cultural Publications, Tehran (1995)
3. Said, H.M.: *Al-Biruni's Book on Mineralogy – The Book Most Comprehensive in Knowledge on Precious Stones*. Pakistan Hijra Council, Islamabad (1989)

4. Kirchner, E.: Origin and spreading of Tusi's ideas on color ordering. Paper presented at the twelfth congress of the International Colour Association (AIC), Newcastle, 8–12 July 2013. Conference proceedings, pp. 1197–1200
5. Sabra, A.I.: The Optics of Ibn al-Haytham. Warburg Institute, London (1989). Two vols
6. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) *Encyclopaedia of Islam*, vol. V, pp. 698–707. Brill, Leiden (1991)
7. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res. Appl.* **40**, 5–16 (2015)

Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, AJLeiden,
The Netherlands

²AUT Textile Department, Amirkabir University
of Technology, Tehran Polytechnic, Tehran, Iran

At the end of the thirteenth century, Kamal al-Din Hasan ibn Ali ibn Hasan al-Farisi studied at the school of Tabriz (Azerbaijan, Iran). His teacher was the famous astronomer Qutb al-Din al-Shirazi, who in turn was a student of Nasir al-Din al-Tusi.

Al-Farisi felt “perplexity” when he found several inconsistencies and errors in the classical works about optics [1]. His teacher then managed to obtain a manuscript copy of the *Kitab al-Manazir* from Ibn al-Haytham, which was brought in “from a very distant land” (probably Egypt). Al-Farisi was greatly impressed by this work and decided to write a detailed commentary on it. He discussed it in great detail and completed it by adding appendices with other optical writings. Al-Farisi also included corrections to these texts, aptly calling the resulting work *Tanqih al-Manazir* (Revision of the [*Kitab*] *al-Manazir*) [2].

In one of al-Farisi's comments in the *Tanqih*, he described how an object is seen in a certain color under sunlight, but in a different color under moonlight, and yet in another color in the light of fire. From this, al-Farisi concluded that colors are not really present in objects, but depend on

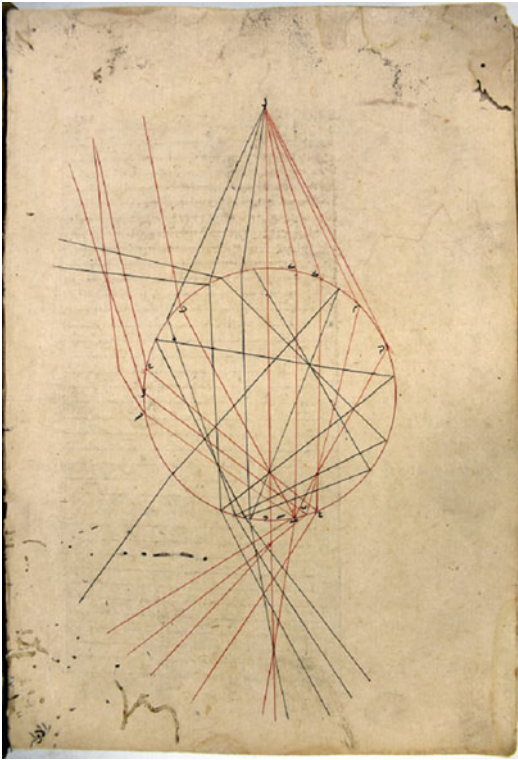
illumination [3]. The role of incident light had therefore been changed from being a mere catalyst for color vision as in the ancient Greek theories to being as prominent as in modern color theory [4].

Together with his teacher Qutb al-Din al-Shirazi, al-Farisi tried to find explanations why the famous experiment of Ptolemy, with a spinning top (or mill stone) having sectors painted in different colors, led to new colors [5].

Regarding the colors of the rainbow, Ibn al-Haytham had supported the Aristotelian idea that these were due to a mixture of light and darkness. Al-Farisi rejected this concept. He argued that if it was true, then the colors of the rainbow would be ordered from bright to dark. Also, the secondary rainbow would then have the same color order. Both are not supported by observation [6].

But al-Farisi would become most famous for his experimental study of the formation of rainbow colors. Inspired by Ibn Sina's work and by Ibn al-Haytham's *Kitab al-Manazir* and the latter's treatise on the burning glass sphere, he filled a glass sphere with water and considered this as a model for a droplet of rain water in the atmosphere. He then studied the resulting reflection and refraction of light in a dark room (see Fig. 1). This led him to the first correct explanation of the colors of the rainbow, which he described in the *Tanqih*. Interestingly in Germany, at approximately the same time also Theodoric of Freiburg, equally inspired by the *Kitab al-Manazir*, carried out the same experiment and formulated the same conclusions.

It was recently discovered that in his *Tanqih*, al-Farisi included the text of al-Tusi on color ordering, in which five paths were proposed to go from white to black [7]. In this way, this text would become available to many later generations. Al-Farisi also sought for an explanation of the different orderings of the colors in the primary and secondary rainbow. His dark room experiments let him conclude that the various colors of the rainbow were produced by a superposition of different images as projected after reflections and refraction in the sphere [8]. Thereby, the colors became a function of the positions and luminous intensities of the composing images [9].



Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan, Fig. 1 Double refraction and single reflection in a raindrop. Simultaneously with Theodoric of Freiberg, al-Farisi was the first to give the correct explanation of the rainbow in his *Tanqih al-Manazir* (©Library of the Masjid Sepahsalar, Tehran)

The *Tanqih al-Manazir* would become widely spread in the Muslim world, where it was used in academic classrooms, and would be commented upon until the sixteenth century. Therefore, in the Muslim world, this would be the main textbook on optics for more than three centuries. Since it would never be translated into any Western language, its influence on European science would remain marginal at best.

References

1. Kheirandesh, E.: Mathematical sciences through Persian sources: the Puzzles of Tusi's optical works. In: Pourjavady, N., Vesel, Z. (eds.) *Sciences, techniques et instruments dans le monde Iranien (Xe – XIXe siècle)*. Presses universitaires d'Iran and Institut Français de recherches en Iran, Téhéran (2004)

2. Kamal al-Din al-Farisi: *Tanqih al-Manazir*, Deccan, Hyderabad (1928) Two volumes
3. Gärtje, H.: Zur Farbenlehre in der muslimischen Philosophie. *Der Islam*, **43**, 280–301 (1967)
4. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res. Appl.* **40**, 5–16 (2015)
5. Vernet, J.: Mathematics, astronomy, optics. In: Schacht, J., Bosworth, C.E. (eds.) *The Legacy of Islam*, p. 483. Clarendon, Oxford (1974)
6. van Campen, M.: *De regenboog bij de Arabieren*. University of Utrecht, Utrecht (1988) MSc Thesis, 95
7. Kirchner, E.: Origin and spreading of Tusi's ideas on color ordering. Paper presented at the twelfth Congress of the International Colour Association (AIC), Newcastle, 8–12 July 2013. Conference proceedings, pp. 1197–1200
8. Rashed, R.: Kamal al Din. In: Gillespie, C.C. (ed.) *Dictionary of Scientific Bibliography*, pp. 212–219. Scribner's Sons, New York (1970)
9. Rashed, R.: Geometrical optics. In: Rashed, R. (ed.) *Encyclopedia of the History of Arabic Science*, vol. 2, pp. 643–671. Routledge, London (1996)

Al-Haytham (Alhazen), Abū 'Alī al-Ḥasan ibn al-Ḥasan ibn

Eric Kirchner

Kirchner Publications, Leiden, The Netherlands

Abu Ali al-Hasan Ibn al-Haytham (Latinized name: Alhazen) made important discoveries in astronomy, mathematics, and optics. His *Kitab al-Manazir* (Book on Optics) made the works of Aristotle, Ptolemy, and Euclid obsolete. He merged their classical theories, combining the mathematical, physical, and physiological aspects into one unified optical theory. The *Kitab al-Manazir* also introduced the distinction between optics and physiology/psychology that is a cornerstone of modern optics and colorimetry.

In his work, Ibn al-Haytham attempted to show how illumination, hue, and saturation combine together into color perception. Although he was not able to establish a complete system of color attributes, his account is largely consistent. “In terms of scope, details and comprehensiveness of treatment, Ibn al-Haytham's observations on the subject of color perception are unequalled in

any single writer before him” (Vol. 2, p. 43 in Ref. [1]).

The *Kitab al-Manazir* would be the dominant text on optics for several centuries to come, both in the East and West. Latin translations appeared from 1200 to 1572 (*Perspectiva, De aspectibus, Opticae Thesaurus*) and in Italian around 1350 (*Prospettiva*). Basing themselves on what they read in this book, Kamal al-Din al-Farisi and Theodoric of Freiburg around 1300 discovered the correct explanation for the colors of the rainbow. It also made Kepler to formulate in 1604 the theory of the retinal image and Rudolph Snel van Royen to discover a few years later what is now known as Snellius’ law of refraction. In the East, Ibn al-Haytham’s work survived through its thorough treatment and further elaboration by Kamal al-Din al-Farisi (d. 1318).

Ibn al-Haytham was born around 965 AD in Basra (Iraq) where he became a famous mathematician. Caliph al-Hakim of Cairo, impressed by Ibn al-Haytham’s claim that he could regulate the flow of the Nile, persuaded him to come to Egypt. However, Ibn al-Haytham soon found out that he was not up to the task. He fell in disgrace and was placed under house arrest for many years. This proved to be beneficent for continuing his studies in mathematics, astronomy, and optics.

According to Aristotle, light was a manifestation of a change of the state of the medium from opaque to transparent. Ptolemy had considered vision to be the result of visual rays, spreading out from the eye that were reflected or refracted by surfaces of objects. Ibn al-Haytham (as well as Ibn Rushd) proposed an alternative view, that light has a much more active role in color vision. It is light which is seen, according to Ibn al-Haytham [2]. Moreover, light is directed toward the eye instead of spreading from it. Ibn al-Haytham was the first to consider light as an entity by itself, traveling from visible objects to the eye, according to mathematical laws originally proposed by Euclid [3]. Regarding the role of the medium for color vision, Ibn al-Haytham adopted the view from al-Kindi (d. 873) that the medium plays a passive role in color vision in the sense that it should not block vision.

In Ibn al-Haytham’s theory, color is a distinct property of material bodies. He stated that color

and light are distinct but that colors behave exactly as light does in transmission, reflectance, and refraction. Therefore, Ibn al-Haytham’s theory of light is at the same time a theory of color [4]. But apart from formulating these theories, Ibn al-Haytham went further and designed many experiments by which he verified the proposal step by step and afterward subjected the results to thorough mathematical analysis [5] (and also Vol. 2, p. xli in Ref. [1]). Nevertheless, Ibn al-Haytham also explored some more philosophical treatments of color and light [5].

Ibn al-Haytham’s view of the active role of light in color vision made him investigate many crucial aspects of color vision. He described how changing the type and intensity of light affects the color appearance of objects and how a stronger light intensity increased perceived color differences [6, 7]. Thus Ibn al-Haytham came close to what has been called a key notion in colorimetry that color is the product of the eye, the light, and the object. In the words of Ibn al-Haytham: “This behavior indicates that the eye observes the colors of colored objects only according to the colors that fall on them” [6, 7].

Ibn al-Haytham was one of the first to use the word saturation as describing an aspect of color. Ibn al-Haytham used this word in his *Kitab al-Manazir*. Generally it refers to the sensation one feels after a copious meal, indeed analogous to the modern English word saturation. In modern Arabic, the same word is still used for color saturation [8]. About this and other words related to color that appear in the *Kitab al-Manazir*, see Vol. 2, p. 43 in Ref. [1].

In other aspects, Ibn al-Haytham followed the classical traditions. He supported the Aristotelian view that the colors of the rainbow are the result of mixing light and darkness and the Ptolemaic interpretation that color mixing in the spinning top is a visual illusion.

References

1. Sabra, A.I.: The Optics of Ibn al-Haytham, 2 vols. Warburg Institute, London (1989)
2. Smith, M.A.: Alhacen’s theory of visual perception. Trans. Am. Philos. Soc. **91**, 4 (2001)

3. Darrigol, O.: The analogy between light and sound in the history of optics from the ancient Greeks to Isaac Newton. *Centaurus*. **52**, 117–155 (2010)
4. Sabra, A.I.: Ibn al-Haytham's revolutionary project in optics: the achievement and the obstacle. In: Hogendijk, J.P., Sabra, A.I. (eds.) *The Enterprise of Science in Islam*, pp. 85–118. MIT Press, Cambridge, MA (2003)
5. el-Bizri, N.: Ibn al Haytham et le problème de la couleur. *Oriens Occidens*. **7**, 201–226 (2009)
6. Wiedemann, E.: Zu Ibn al Haithams Optik. *Arch. Gesch. Naturwiss. Techn.* **3**, 1–53 (1910–1912)
7. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Color Res. Appl.* **40**, 5–16 (2015)
8. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) *Encyclopedia of Islam*. Brill, Leiden (1991)

Alizarin

Tuane Cristina dos Santos
North Carolina State University, Raleigh,
NC, USA

Synonyms

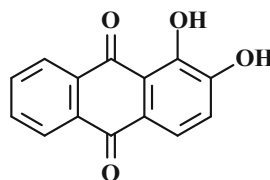
1,2-Dihydroxy-9,10-anthracenedione; 1,2-Dihydroxyanthraquinone; Alizarin red; Alizarine; CI Mordant red 11; Madder lake; Turkey red

Definition

Alizarin is a redish-orange anthraquinone dye with formula $C_{14}H_8O_4$. It can be extracted from the roots of plants of the madder genus or be obtained synthetically. Historically, it has been used for dyeing textiles.

Overview

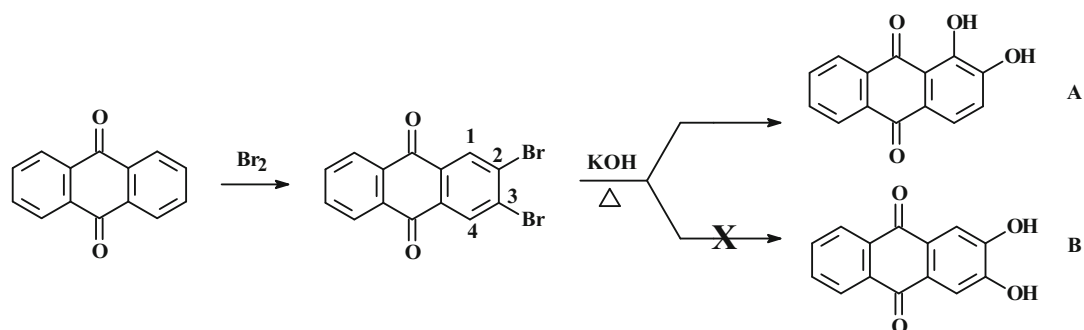
Alizarin (1,2-dihydroxyanthraquinone) is normally the main colorant extracted from the dried roots of the plant *Rubia tinctorum* L (Madder) and is undoubtedly the best known of the anthraquinone natural dyes. Its chemical formula is shown in Fig. 1. Madder has been used as a dye source for many centuries, being cultivated in central



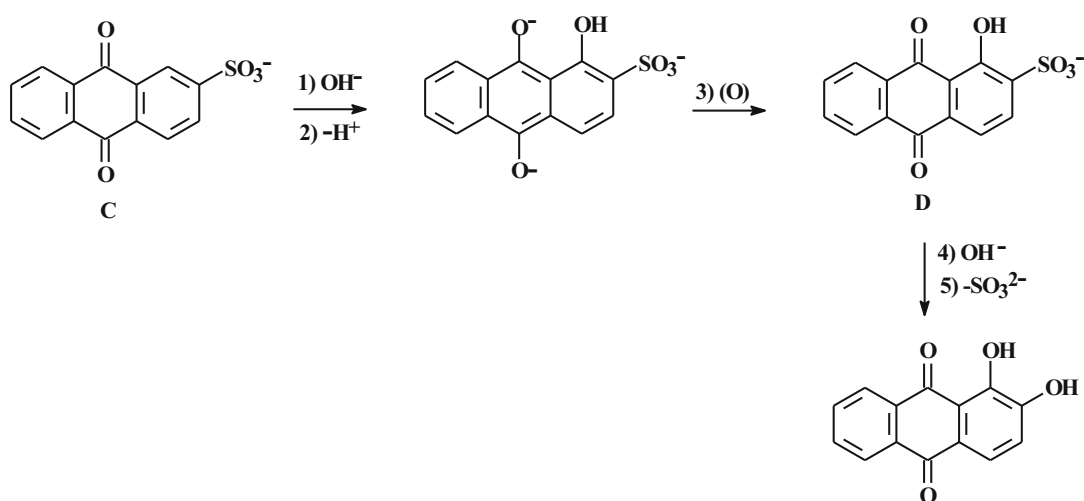
Alizarin, Fig. 1 Structure of alizarin

Asia and Egypt, where it was grown as early as 1500 BC. In 1826, alizarin was first isolated from madder roots by the chemist Pierre-Jean Robiquet who found two colorants in the madder root extract: the red dye alizarin (~4% by weight) and the more rapidly fading dye purpurin [1]. The 1865 discovery of the chemical constitution of benzene opened the door to the elucidation of the molecular structure of alizarin by Graebe and Liebermann in 1867 and to synthetic approaches to this interesting natural colorant. Thus, anthraquinone dyes as a family of industrial compounds can trace their roots back to the determination of alizarin's structure.

In 1868, alizarin became the first natural dye component produced synthetically [2]. Its initial synthesis involved a two-step sequence from anthraquinone, namely bromination and hydroxylation, as illustrated in Fig. 2. There it can be seen that bromination occurs in two of the beta positions (2,3-positions) but fusion with alkali in the second step produces 1,2-dihydroxyanthraquinone (alizarin; **A**) rather than the 2,3-dihydroxyanthraquinone isomer (**B**) [3]. The cost of bromine and practical difficulties associated with scale-up prevented commercialization of this approach. It was later determined that the synthesis of alizarin from anthraquinone-2-sulfonic acid was commercially viable and the method of choice. The principal synthesis route involved the conversion of anthraquinone-2-sodium sulfonate ("silver salt") to 2-hydroxyanthraquinone using hot KOH and oxidation with sodium nitrate or sodium chlorate in alkaline media to give alizarin [4]. Thus, this process not only gave the expected replacement of SO_3Na by OH but also introduced an OH group in the 1-position [5]. An overview of a proposed reaction mechanism for the series of addition and elimination steps leading to alizarin is given



Alizarin, Fig. 2 Early synthesis of alizarin



Alizarin, Fig. 3 Steps in the commercial synthesis of alizarin

in Fig. 3. Step 1 starts with anthraquinone-2-sulfonate (C) and is the first of two hydroxide (OH^-) addition steps. This is followed by the loss of a proton (Step 2) and air oxidation (Step 3), to give 1-hydroxyanthraquinone-2-sulfonate (D). The second addition of OH^- (Step 4) is accompanied by elimination of sulfite (SO_3^{2-}) (Step 5) to give alizarin [6, 7]. Synthetic alizarin became very appealing, as it was cheaper and did not require the laborious extraction of the colorant from the plant material [8]. These advantages virtually eliminated the market for natural madder, ending the presence of huge cultivations in parts of Europe that dated back to 1800. Subsequently, however, natural madder could be obtained from Africa and South America [9].

Alizarin is slightly soluble in water but readily soluble in alcohol, ether, and alkaline solutions. Natural alizarin can be used as a pH indicator, as its color varies from scarlet to pink to red with a bluish tint depending on the strength of alkali or acid used. Strong alkali produces a violet-blue color, while a strong acid gives yellowish-red [1].

Other applications include its use as a food colorant, in cosmetics, and in textile dyeing [10]. Regarding its textile use, alizarin is classified as a mordant dye (CI Mordant Red 11), since it readily forms a dye-metal complex on fibers, notably when using aluminum ions (Al^{3+}). Its complexes show more intense colors and an enhanced set of fastness properties (stability under end-use conditions) compared with the parent (uncomplexed)

dye. In the actual dyeing process, mordant and dye combine to form a less soluble colorant within the textile fibers, which makes the dyed fiber resistant to external influences such as washing and finishing processes [11]. Furthermore, the choice of mordants used influences the final shade obtained. For example, pink can be obtained when alizarin complexes with tin, violet when it complexes with iron, and red upon complexing with aluminum, especially on leather and cotton [12]. In the brown trivalent chromium–alizarin complex, the chromium ion is coordinated with three bidentate alizarin molecules [13]. In turn, this complex bonds to protein-based fibers (wool and silk) through the formation of hydrogen bonds between the polypeptide linkages and the dye [9]. Nowadays, alizarin is mostly used as a staining agent in biochemical assays and in water-color and oil painting.

Although natural dyes are generally considered to be eco-friendly, since they are mainly derived from plants and other renewable sources, their production, mode of application to fibers (use of metallic mordants), mode of action in the human body, and environmental fate are not completely free from concerns [14]. With this in mind, studies concerning the carcinogenic and mutagenic potential of madder extract and alizarin have been conducted. Madder-related studies performed in rats showed chronic toxicity targeting kidney, liver, and blood cells [15], and both renal tubule cell and hepatic carcinogenicity [16]. Moreover, madder was mutagenic to strains TA100 and TA98 of *Salmonella typhimurium*, with the observed mutagenicity attributed to components present in the madder root extracts other than alizarin [17]. However, in another study, alizarin was strongly mutagenic to strain TA2637 with metabolic activation [18].

In medicinal studies, natural alizarin has been shown to have antitumor properties (in colon cancer cells) and low dosages exhibited selective inhibition of osteosarcoma and breast carcinoma cell proliferation in in vitro studies [19]. Alizarin has also exhibited antifungal properties, with the potential to be used in the treatment of candidiasis [20].

Cross-References

- [Coloration, Fastness](#)
- [Coloration, Mordants](#)

References

1. De Santis, D., Moresi, M.: Production of alizarin extracts from *Rubia tinctorum* and assessment of their dyeing properties. *Ind. Crop. Prod.* **26**(2), 151–162 (2007)
2. Singh, H.B., Bharati, K.A.: Enumeration of dyes. In: Singh, H.B., Bharati, K.A. (eds.) *Handbook of natural dyes and pigments*, p. 33. Woodhead Publishing India, New Delhi (2014)
3. The Complete Technology Book on Dyes & Dye Intermediates, National Institute of Industrial Research, Delhi – 7, India, 2003, p. 416
4. Perkin, W.H.: The History of Alizarin and Allied Colouring Matters, and Their Production from Coal Tar: Two Lectures Delivered Before the Chemical Section of the Society of Arts, May 8th and 22nd, 1879. W. Trousce (1879)
5. Puntener, A.G., Schlesinger, U.: Natural dyes. In: Freeman, H.S., Peters, K.A. (eds.) *Colorants for Non-textile Applications*, p. 424. Elsevier (2000)
6. Hallas, G.: Chemistry of anthraquinoid, polycyclic and miscellaneous colorants. In: Shore, J. (ed.) *Colorants and Auxiliaries*, vol 1 – Colorants, p. 235. Society of Dyers and Colourists, Bradford (1990)
7. Zollinger, H.: *Color Chemistry*, p. 271. Wiley-VCH, Zürich (2003)
8. Vankar, P.S.: Chemistry of natural dyes. *Resonance*. **5**(10), 73–80 (2000)
9. Angelini, L.G., Pistelli, L., Belloni, P., Bertoli, A., Panconesi, S.: *Rubia tinctorum* a source of natural dyes: agronomic evaluation, quantitative analysis of alizarin and industrial assays. *Ind. Crop. Prod.* **6**(3–4), 303–311 (1997)
10. Bechtold, T., Mussak, R.: *Handbook of Natural Colorants*. Wiley (2009)
11. Waheed, S., Alam, A.: Studies of some natural dyes. *J. Chem. Soc. Pak.* **26**, 255–263 (2004)
12. Ferreira, E.S., Hulme, A.N., McNab, H., Quye, A.: The natural constituents of historical textile dyes. *Chem. Soc. Rev.* **33**(6), 329–336 (2004)
13. Jones, F.: The chemistry and properties of metal-complex dyes. In: Shore, J. (ed.) *Colorants and Auxiliaries*, vol 1 – Colorants, p. 205. Society of Dyers and Colourists, Bradford (1990)
14. Agnhage, T., Perwuelz, A., Behary, N.: Towards sustainable *Rubia tinctorum* L. dyeing of woven fabric: how life cycle assessment can contribute. *J. Clean. Prod.* **141**, 1221–1230 (2017)

15. Inoue, K., Shibutani, M., Masutomi, N., Toyoda, K., Takagi, H., Takahashi, M., et al.: One-year chronic toxicity of madder color in F344 rats—induction of preneoplastic/neoplastic lesions in the kidney and liver. *Food Chem. Toxicol.* **46**(10), 3303–3310 (2008)
16. Inoue, K., Yoshida, M., Takahashi, M., Shibutani, M., Takagi, H., Hirose, M., et al.: Induction of kidney and liver cancers by the natural food additive madder color in a two-year rat carcinogenicity study. *Food Chem. Toxicol.* **47**(1), 184–191 (2009)
17. Jäger, I., Hafner, C., Welsch, C., Schneider, K., Iznaguen, H., Westendorf, J.: The mutagenic potential of madder root in dyeing processes in the textile industry. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **605**(1), 22–29 (2006)
18. Li, C., Li, J., Huang, G., Li, Q., Yan, L.: Assessment of potential toxicity of alizarin through resonance light scattering. *Asian J. Chem.* **25**(2), 1015 (2013)
19. Fotia, C., Avnet, S., Granchi, D., Baldini, N.: The natural compound Alizarin as an osteotropic drug for the treatment of bone tumors. *J. Orthop. Res.* **30**(9), 1486–1492 (2012)
20. Manoharan, R.K., Lee, J., Kim, Y., Lee, J.: Alizarin and chrysazin inhibit biofilm and hyphal formation by *Candida albicans*. *Front. Cell. Infect. Microbiol.* **7**, 447 (2017)

Alizarin Red

► [Alizarin](#)

Alizarine

► [Alizarin](#)

Al-Tusi, Nasir al-Din

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

Muhammad ibn Muhammad ibn Hasan al-Tusi, usually known as Nasir al-Din al-Tusi, was born in Tus (Iran) and later worked in Maragha (Azerbaijani) and Baghdad (Iraq). His influence

reaches into many fields [1]. His work on reforming Ptolemaic theoretical astronomy would be crucial for later astronomers, including Copernicus. In mathematics he published landmark editions of the works of Euclid and Archimedes and developed trigonometry as a discipline separated from astronomy. He wrote several works on optics [2]. For Shiite theology, al-Tusi wrote an important work on ethics, and he also authored the first systematic treatment of rationalist theology in 12 imam Shiism, a work still central in Shiite theological education.

Al-Tusi was born in a Twelver Shi'a family. Under patrons at Ismaili courts, he became a famous mathematician. When in 1256 under Mongol rule, Hulagu destroyed the Abbasid Empire, al-Tusi had taken refuge at the last Ismaili stronghold located on the mount of Alamut. After the Mongols destroyed this mountain fortress, Hulagu personally saved al-Tusi's life since the ruler's interest in astrology made him respect al-Tusi's astronomical knowledge. Al-Tusi persuaded Hulagu to support building the first full-scale astronomical observatory in the world at Maragha. With al-Tusi as director, it would collect a mass of observation data during about 50 years, and it would inspire later observatories in Samarkand and India and possibly even Tycho Brahe's observatory in Denmark.

Al-Tusi thoroughly studied the works of Ibn Sina. When after teaching Ibn Sina's color theory one of his students wanted to know more about it, al-Tusi replied in a letter:

Regarding the production of colors from black and white there are numerous paths, from which one gradually walks from white to black. The path through yellow belongs there: First by the mixing of dense and fire, both in small amount, the straw-yellow is produced, then the lemon-yellow, then the saffron-yellow, then the orange-yellow, then the grenade-yellow, then in it the tendency towards black increases, according to the increase in the number of dense particles and the decrease of fire, until it becomes black.

Another path goes through red. First it becomes rosy, then like evening-red, then blood-colored, then purple, then violet, violet-colored. [...]

This all occurs according to the differences of particles in transparency, opacity (density), light and darkness. Now and then one sees a color together with another, and a different color is

produced, such as green from yellow and blue, verdigris from green and white. There are infinitely many of such arrangements, and some are often found in small particles of plants and animals. Anyone who observes them is surprised by their number. [3, 4]

Thus, while Ibn Sina had specified three paths from white to black, al-Tusi described five of such paths. They go via yellow, red, green, blue, and gray, *as illustrated in the Figure above*. Remarkably Forsius in 1611 also described five different tint/shade scales from white to black. This may all be compared to the German monk Theophilus (ca. 1120), who described how to produce up to 12 grades in a tint/shade scale. The text of al-Tusi shows that he must have considered color space to be two dimensional. Interestingly, at approximately the same time in Latin Europe, Grosseteste had argued on theoretical grounds that color space is three dimensional [5, 6].

Also in his work on minerals and gem stones, the *Tansukhname-yi Ilkhani* (the book on precious stones for the Ilkhan [i.e., for Mongol ruler Hulagu]), al-Tusi describes a color theory [3, 7]. Written primarily by Nishaburi, and largely copied by al-Tusi and Kashani, these texts are the first to describe a limited hue scale. It describes that by mixing blue and yellow in different proportions, colors are produced that change gradually from blue, via green, to yellow. This description represents a great step forward from the Aristotelian point of view that stated that green is one of the colors painters cannot produce. Although several scholars before Nishaburi and al-Tusi had mentioned that green can be produced by mixing blue and yellow, no earlier scholar had described that depending on the mixing ratios, different hues of green are produced [6].

The common opinion among scholars since Aristotle had been that by mixing black and white all colors can be produced. But following Nishaburi, Tusi wrote that “if white color and black color are mixed with each other, an incense-grey color will result.” This had been stated only twice before in history and in much less clear wordings. Clearly this statement, which was a starting point for Newton’s optical work, was not made for the first time by Scaliger in 1557, as is generally thought [8].

References

1. Nasr, S.H.: al-Tusi. In: Gillespie, C.C. (ed.) Dictionary of Scientific Bibliography, vol. 13, pp. 508–514. Scribner’s Sons, New York (1970)
2. Kheirandesh, E.: Mathematical sciences through Persian sources: the puzzles of Tusi’s optical works. In: Pourjavady, N., Vesel, Z. (eds.) Sciences, techniques et instruments dans le monde Iranien (Xe – XIXe siècle). Presses universitaires d’Iran and Institut Français de recherches en Iran, Téhéran (2004)
3. Kirchner, E., Bagheri, M.: Color theory in medieval lapidaries: Nishaburi, Tusi and Kashani. *Centaurus*. **55**, 1–19 (2013)
4. Persian edition in Nourani, A.: Ajvebat al-Masā’el al-Nasīrīyah. Institute for Humanities and Cultural Studies, Tehran (2005). German translation in: Wiedemann, E.: Über die Entstehung der Farben nach Nasir al Din al Tusi. *Jahrbuch der Photographie* 22, 86–91 (1908). Reprinted in Wiedemann, E.: Aufsätze zur arabischen Wissenschaftsgeschichte, pp. 256–261. Georg Olms, Hildesheim (1970)
5. Smithson, H., Dinkova-Bruun, G., Gasper, G.E.M., Huxtable, M., McLeish, T.C.B., Panti, C.: A three-dimensional color space from the 13th century. *J. Opt. Soc. Am. A*. **29**, A346–A352 (2012)
6. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res.* **40**, 5–16 (2015)
7. Razavi, M.: *Tansukhname-ye Ilkhany*. Miras-i Maktub, Tehran (1969)
8. Shapiro, A.E.: Artists’ colors and Newton’s colors. *Isis*. **85**, 600–630 (1994)

Anchoring Theory of Lightness

Alan Gilchrist

Department of Psychology, Rutgers University,
Newark, NJ, USA

Synonyms

Achromatic color; Perceived reflectance; Perceived value

Definition

Lightness refers to the white, black, or gray shade of a surface. The basic problem is that the light reaching the eye from a given surface, as the product of surface reflectance and illumination,

does not specify the reflectance of the surface. A black in sunlight can reflect more light than a white in shadow. No computer program exists that can identify the reflectance of an object in a photo or video. Any possible solution must exploit the context surrounding the object. Anchoring theory proposes to solve the problem by grouping patches of the retinal image into regions of equal illumination, called frameworks, then computing lightness values within each framework using rules of anchoring combined with luminance ratios.

Background

Lightness refers to the perceived shade of white, gray, or black of a surface and is sometimes called perceived reflectance. The reflectance of a surface is the percentage of light it reflects. White surfaces reflect about 90% of the light that illuminates them while black surfaces reflect only about 3%. Given the light-sensitive receptor cells in the eye, one might think that the perception of lightness could be easily explained. But the amount of light reflected by a surface, called luminance, depends not only on its reflectance but also on the intensity of light illuminating it. And the intensity of illumination varies from place to place and time to time. Thus, a black surface in sunlight can easily reflect more light than a white in shadow. In fact, any shade of gray can reflect any amount of light.

The ability to perceive the correct shade of gray of a surface despite variations (both spatial and temporal) in illumination is called lightness constancy, or more specifically, illumination-independent lightness constancy. The ability to perceive the correct shade of gray despite variations in the luminance of the background surrounding a target surface is called background-independent constancy. Neither type of constancy is complete. Variations in illumination level produce failures of constancy as described in the traditional literature, while variations in background produce failures of constancy that have been called contrast effects.

Summary of the Theory

According to anchoring theory [1], the lightness of a target surface is codetermined by its relationship to the field of illumination in which it is embedded, called the local framework, and its relationship to the entire visual field, or global framework. The values computed within these local and global frameworks are combined in a weighted average. Within each framework, the value is based on the ratio between the luminance of the target surface and the highest luminance in the framework, called the anchor. The weight of a framework depends on its size and complexity. Framework boundaries include corners, occlusion contours, and penumbrae.

Frameworks

Helmholtz [2] suggested that given the luminance of a target surface, the visual system would have to somehow take into account the intensity of light illuminating that surface in order to compute its reflectance. But in fact this is overkill. To compute lightness, the visual system needs to know only which surfaces are under *the same level of illumination*. In anchoring theory, each scene is perceptually segmented into regions of homogeneous illumination, as Koffka [3] suggested. Thus any luminance differences within each of these frames of reference must represent reflectance differences. By comparing only surfaces that are equally illuminated, reflectance and illumination are disentangled, not by estimating illumination but by holding it constant.

Frameworks are equivalent to perceptual groups, in which retinal patches are grouped by common illumination, a process that parallels the more traditional gestalt idea of grouping by which objects are segregated, which can be thought of as grouping by common reflectance.

Anchoring

In 1948, Wallach [4] proposed the idea that lightness depends simply on the ratio between the

luminance of a surface and the luminance of the surrounding region. There is clearly much validity in this proposal. A homogenous disk of constant luminance can be made to appear as any shade of gray from white to black simply by varying the luminance of a homogeneous ring that surrounds it. However, relative luminance per se can only produce relative shades of gray. For example, consider two adjacent surfaces whose luminance values stand in a 5:1 ratio. These could be white and middle gray. But they might also be light gray and dark gray, middle gray and black, or any number of reflectance pairs. To determine the specific shade of a surface requires, in addition to relative luminance, an anchoring rule – that is, a rule that links some value of relative luminance to some value on the gray scale. Anchoring theory invokes a highest luminance rule according to which the highest luminance within a region of illumination is perceived as white, and this value serves as the standard to which lower luminance values are compared.

Simple and Complex Images

Anchoring theory makes an important distinction between simple images and complex images. A simple image is one in which a single level of illumination fills the entire visual field while a complex image, typical of most images we see, contains multiple levels of illumination. The theory also describes how the rules for simple and complex images are related to each other.

Simple Images

In a simple image, the formula used to predict lightness value of a given target surface is **Lightness** = $T/H * 0.9$, where **T** is the luminance of the target, **H** is the highest luminance in the image, and **0.9** is the reflectance of white.

Geometric relationships, in addition to photometric relationships, also play a role in anchoring. Lightness depends, to a limited extent, on the relative area of surfaces. In general, lightness increases as relative area increases. But according

to anchoring theory, this effect of “the larger the lighter” applies only when that darker surface fills more than half of the area of the visual field.

Complex Images

While it is not hard to exhaust the rules by which lightness is computed in simple images, the more important goal of a theory of lightness is to explain how lightness is computed in complex images typical of the real world. Anchoring theory makes the claim that complex images are related to simple images according to two principles. The first is that the same rules of lightness computation found in simple images apply to frameworks embedded in complex images. The second involves a kind of crosstalk between frameworks called codetermination.

Codetermination

The term *codetermination* was coined by Kardos [5], a gestalt psychologist from Budapest. He proposed a friendly amendment to Koffka’s idea of frameworks, observing that the lightness of a target is not computed exclusively within its framework but is influenced by neighboring frameworks as well. Kardos argued that the lightness of a surface is partly determined by its relative luminance in the field of illumination to which it belongs, called the relevant field, and partly by its relationship to a neighboring field of illumination, which he called the foreign field.

Anchoring theory invokes the closely related concepts of local and global frameworks. The global framework is the entire visual field. Unlike relevant and foreign frameworks, which are exclusive, local and global frameworks are hierarchical. The local framework is nested within the global.

Segmentation and Grouping

Frameworks are the products of both segmentation and its inverse: grouping. Segmentation

factors include cast illumination edges, such as the border of a shadow or spotlight, corners, where the planarity of a surface changes, and occlusion boundaries, where one object partially obscures a more distant object. Such illumination boundaries effectively group retinal patches by illumination level.

Additional gestalt grouping factors are held to create weak frameworks. In the simultaneous lightness contrast illusion, for example, a gray patch on a white background appears slightly darker than the same gray patch on a black background. Although the black and white backgrounds are equally illuminated (and thus the two patches should have the same lightness value), each patch is perceptually grouped with its background due to adjacency and surroundedness and these weak frameworks cause the (weak) illusion.

Veridicality Versus Error

In general, theories of lightness have sought to account for lightness constancy, and the degree to which perceived lightness corresponds to physical reflectance. However, an adequate theory of lightness must also explain failures of constancy. The Kardos theory of codetermination was the first lightness theory to confront this problem. The failure of constancy within a relevant field is due to the unwanted influence of the foreign field. Anchoring theory endorses this construction but also broadens it to include a larger class of what can be called lightness errors. This class includes lightness illusions in addition to failures of constancy. Illusions and constancy failures have traditionally been treated separately but both constitute errors.

The most basic and familiar illusion of lightness is called simultaneous contrast. Two identical gray patches are placed, respectively, on adjacent black and white backgrounds. The patch on the black background appears lighter than the patch on the white background. The application of anchoring theory to this illusion is rather simple. The illusion is obviously composed of two frameworks, corresponding to the two backgrounds.

These two frameworks taken together can be considered the global framework. For each gray patch, both a local and a global value are computed. In the global framework, both gray patches are computed to be middle gray. But they get different values when computed locally. The patch on the black background is given a value of white in its local framework, because it is the highest luminance. However, the patch on the white background is computed to be middle gray, relative to its white surround. When the local and global values for each patch are combined in a weighted average, the patch on the black background ends up with a slightly higher value. According to anchoring theory, this “contrast effect” is relatively weak because the two local frameworks are weak, due to their lack of articulation.

According to anchoring theory, the simultaneous contrast illusion is not caused by lateral inhibition or any exaggeration of edge differences but rather by the fact that one patch is perceptually grouped with the black background while the other patch is grouped with the white background. Consistent with this grouping approach, several authors have produced reverse contrast illusions. These illusions are variations of simultaneous contrast in which a gray patch on the black background appears darker, not lighter, than an identical gray patch on the white background. This reversed effect is created by making the gray patch a member of a group of patches that are opposite in lightness from the black background. Thus, the gray patch on the black background is part of a matrix of white patches of the same size and shape. The gray patch on the white background is likewise part of a group of black patches. These reverse contrast illusions support the concept of grouping that is fundamental to gestalt theories in general and anchoring theory in particular.

Weighting

To implement the concept of codetermination, a provisional lightness value is computed for the target within each framework, including the

global, and then a weighted average of these values is computed. The weight of each framework in this average depends on the size of the framework and the degree of complexity within it.

The factors of framework size and articulation come from the early work of David Katz [6]. Having invented some of the fundamental methods for studying lightness, Katz conducted experiments testing the factors that lead to strong and weak constancy. He found that the degree of constancy varies in proportion to the size of a framework, which he called field size, and even more-so to its level of articulation, operationally defined simply as the number of distinct elements within it.

The concepts of field size and articulation are given a slightly different interpretation in anchoring theory. Rather than factors consistently leading to good constancy, they are treated as factors that determine the relative weight of a framework. Thus, for example, replacing the white and black backgrounds of the simultaneous contrast illusion by collages of light gray and dark gray patches, respectively, increases the weight of the two backgrounds and increases the size of the illusion. Constancy, that is, is reduced, not increased.

The general formula for a target surface belonging to the global framework and one local framework is:

$$\text{Lightness} = W(T/H_l * 0.9) + (1 - W_l) \times (T/H_g * 0.9).$$

in which **W** is the weight of the local framework, **T** is target luminance, **H_l** is the highest luminance in the local framework, **H_g** is the highest luminance in the global framework, and **0.9** is the reflectance of white.

Staircase Gelb Effect

Anchoring theory was inspired by an empirical finding called the staircase Gelb effect. Gelb had shown that a piece of black paper appears white when suspended in midair and illuminated by a spotlight. In the staircase version, a row of five squares arranged in an ascending series from

black to white is placed within the spotlight. This produces a dramatic compression of the perceived range of grays. Although the squares range from white to black, they appear to range only from white to a light middle gray, a phenomenon now known as gamut compression.

This finding seriously undermined several earlier decomposition theories of lightness but suggested an interaction between the bright illumination on the five squares and the dimmer illumination of the room. According to anchoring theory, each of the squares is assigned its correct value within the local spotlight framework. But relative to the surrounding room, as long as the spotlight is at least 30 times brighter than the illumination level in the room, each square is assigned the same (global) value of white, given that its luminance is as high as, or higher than, a white surface in the room light. The perceived compression of the lightness range is explained by these equal values assigned to the squares relative to the room illumination.

The Scaling Problem

The mapping of luminance values onto lightness values within a given framework involves both anchoring and scaling. If anchoring concerns which *value* of relative luminance is linked to which *value* of lightness, scaling concerns which *range* of luminance values is linked to which *range* of lightness values. Much as a normal distribution can be characterized by both a measure of central tendency and a measure of dispersion, so a set of luminance values within a framework can be converted to lightness values using an anchoring rule and a scaling coefficient (although the anchor is located at the top, rather than in the center, of the range).

Wallach's ratio principle assumes a one-to-one scaling coefficient. That is, if the luminance ratio between two adjacent surfaces is 3:1, then the ratio of their two perceived reflectance values must also be 3:1. This scaling metric is implicit in the formula given above: **Lightness = T/H * 0.9**. It can be considered the default scaling metric

and applies to large, well-articulated frameworks typical of everyday scenes.

In the staircase Gelb illusion, the range of perceived values in the spotlight is compressed due to the presence of the lower illumination surrounding the spotlight. However, within each framework, the perceived range can be either compressed or expanded, based on its actual range. The theory includes a scale normalization principle, according to which the range of perceived lightness values tends toward the canonical 30:1 white–black range (from 90% to 3%). When the luminance range within a framework is less than 30:1, the lightness range is expanded, and when the luminance range exceeds 30:1, the lightness range is compressed. The degree of expansion or compression is proportional to the deviation of the range from 30:1, but the normalization is only partial.

Anchoring in Other Domains

The anchoring problem arises in other perceptual domains including perceived size and perceived motion. Motion is a good example. There is a good deal of evidence that perceived motion depends on relative motion, but to achieve specific values of motion, relative motion must be anchored. Here, the metaphor of an anchor is even more intuitive than in lightness. In a matrix of relative motions, which elements will be perceived as stationary? In his brilliant writing on perceived motion, Duncker [7] stressed the importance of frames of reference. And he talked specifically about anchoring rules in a way that appears highly consistent with the anchoring theory of lightness.

Strengths and Weaknesses of the Theory

Anchoring theory has wide applicability within lightness perception. It is consistent with an extensive range of empirical data. In general, it accounts for veridical perception to the degree

that perceived lightness is veridical. But unlike most other theories, it also accounts for a wide range of known lightness errors, that is, lightness illusions and failures of constancy. This is important because the pattern of errors shown by humans in lightness perception must be the signature of our visual software.

But anchoring theory remains a work in progress. Recent empirical work appears to require several modifications of the theory, including these:

1. The concepts of nested local and global frameworks should be replaced by the Kardos concepts of relevant and foreign frameworks.
2. New work suggests that the concept of code-termination is implemented, not by averaging the local and global values for each square separately, but by the assignment of an anchor and a scaling coefficient to each framework. Both the anchor and the scaling coefficient are codetermined by the foreign framework and the relevant framework.

These changes provide a more satisfactory account of the staircase Gelb illusion. For a framework of one illumination level surrounded by a region of different illumination, the degree of gamut compression has been shown to correlate with the ratio between the highest luminance values in the two frameworks. But these changes complicate the explanation of simultaneous lightness contrast, for which the original version of the theory was very effective.

Cross-References

- ▶ [CIE Whiteness](#)
- ▶ [Color Constancy](#)
- ▶ [Color Contrast](#)
- ▶ [Ganzfeld](#)
- ▶ [Helmholtz, Hermann Ludwig von](#)
- ▶ [Katz, David](#)
- ▶ [Perceptual Grouping and Color](#)
- ▶ [Simultaneous Color Contrast](#)

References

1. Gilchrist, A.: Seeing Black and White. Oxford University Press, New York (2006)
2. von Helmholtz, H.: Helmholtz's Treatise on Physiological Optics. Optical Society of America, New York (1866/1924)
3. Koffka, K.: Principles of Gestalt Psychology. Harcourt, Brace, and World, New York (1935)
4. Wallach, H.: Brightness constancy and the nature of achromatic colors. *J. Exp. Psychol.* **38**, 310–324 (1948)
5. Kardos, L.: Ding und Schatten [object and shadow]. *Z. Psychol.* **23**, 1–184 (1934)
6. Katz, D.: The World of Colour. Kegan Paul, Trench, Trubner, London (1935)
7. Duncker, K.: Induced motion. In: Ellis, W.D. (ed.) *A Source Book of Gestalt Psychology*, pp. 161–172. Humanities Press, New York (1950)

Ancient Color Naming

► Ancient Color Terminology

Ancient Color Terminology

David Alan Warburton
Institute for the History of Ancient Civilizations,
Northeast Normal University, Changchun, China

Synonyms

[Ancient color naming](#); [Early color lexicons](#); [Early expression of color in language](#)

Definition

Color terms or partitions of color denotata evidenced in ancient language artifacts.

Words and Hues, Languages and Time: An Overview

The sources for understanding the earliest color terms and categories are from the lands around the

eastern Mediterranean Sea, beginning some 5000 years ago. Evidence of color categories from Proto-Cuneiform, Sumerian, Egyptian, and Akkadian (i.e., Babylonian) in Mesopotamia and Egypt (from the end of the fourth millennium BC onward) is followed (from the second half of the second millennium onward) by (Mycenaean) Greek in the West and (Archaic) Chinese in the East. Mesoamerican Maya can be traced back to roots in the first millennium BC, but color is only attested from the first and second millennia AD. Egyptian is the only language which can be traced back from the third millennium BC down to the first millennium AD, meaning that it is the longest lasting of the early languages – and reveals major changes. Maya is much later and the terminology is historically consistent over a long period. As a group, the words discussed here represent six or seven different linguistic families (Sumerian, Afro-Asiatic, Semitic, Indo-European, Sino-Tibetan, and Mayan; the affiliation of Proto-Cuneiform is disputed), with linguistic terms relating to some fundamental colors (dark/black, bright/white, red) with other terms (reds, greens, blues, and yellows) present to some degree in most of these languages.

What is known about the earliest color categories is derived from artifacts and texts. The use of color can be traced back more 300,000 years, but the origins of color vocabulary probably lie in the period since roughly 8000 BC (=10,000 years ago); the earliest texts (from ca. 3500 BC in Mesopotamia) appear millennia later, with color only appearing in a relevant fashion after 3000 BC. By comparison with the languages discussed here, virtually all other languages with complex color vocabularies are either much younger (e.g., Hebrew, Latin), or contemporary (e.g., Eblaitic, Hittite, Ugaritic), and all are linguistically related to the languages discussed here.

Vocabularies in the earliest preserved languages offer representative and definitive evidence concerning the origins of color categorization and its linguistic expression, as well as allowing evaluations of different steps in the process of abstraction and the early linguistic partitioning of perceptual color space.

(For linguistic and historical details, see [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 21]). Before Plato – in Classical Greece (which in the mid-first millennium BC is relatively late for the discourse in this entry) – no thinkers left traces of discussions about color, and thus all of our interpretations about colors from the second and third millennia BC are based on modern interpretations of evidence and not ancient commentaries.

Color Terminology

Black and White, Bright and Dark

The earliest color lexicons from languages of the Middle East and the eastern Mediterranean (Proto-Cuneiform, Sumerian, Akkadian, Egyptian, and early Greek) have words signifying “dark” and “light”; some of these terms also denote something closer to “black” or “white”; and some are related to materials. Egyptian and Sumerian “white” are linked to silver, with the Sumerian also related to “purity,” suggesting that white as a distinct color was added to color vocabulary in historical times. There were usually several terms, as, e.g., Mycenaean Greek *re-u-ko* (the antecedent of Classical Greek *leukós*, “bright,” “white”) was accompanied by *pa-ra-ko*, apparently a white pigment used for dying (Classical *barakis*) – but also used figuratively for a color. Objectively, we can simply confirm that, like most of the more specific terms for “black” in most ancient languages, even Classical Greek *mélas* had a semantic range including “black” and “dark” [16] that encompassed some regions described in English as “brown” – which Schenkel also suggests for ancient Egyptian “black,” *km* (in [12]), and this can be assumed for Sumerian *gege* (meaning “black” and “dark” in Sumerian, and probably appearing in Egyptian as *kk.w* a loan-word meaning only “[extreme] darkness,” “blackness”).

While words for “white” could include “bright,” and those for “black” could include “dark,” these words remained distinct from other words meaning vaguely “bright” or “dark.”

Red

All of the ancient languages have words for some color(s) in the red range. Red ocher was also the first pigment used by humans, hundreds of thousands of years ago, but the word is probably unknown. Unless, the continuity discussed below – from Akkadian *ruššu* “red” to Italian *rosso* – might mean that the English word “red” (like German *rot*) is a direct successor of the earliest word for red, invented relatively recently, but still thousands of years ago (and perhaps before “white”). The Mycenaean Greek Linear B *e-ru-to-ro* is doubtless the same word, and leads into Classical Greek *eruthrós*. Yet *eruthrós* was not the primary word for “red” in second millennium Greek, as there were others, the most important being *po-ni-ko-ro*, a loanword, later *phoîniks*, related to “purple,” “crimson,” and “Phoenician” (parallel to the much rarer Mycenaean *po-pu-re-ja*, later *porphúreos*, a different loan word, used in Classical Greek). Likewise, the most important “red” in Akkadian was not *ruššu*, but *sāmu*, related to *sāmtu*, “red carnelian.” Schenkel has argued for *dšr*, *tjms*, and *mroš* all designating different “reds” in Egyptian. Chinese *hóng* was a pink before becoming a “red,” replacing *chì* (which was earlier red). Red was clearly identified as a color range by the third millennium, but the terminology was unstable. All ancient languages exhibit a very clear tendency for a division of the reddish continuum of color experiences into several different terms related to hues, with loanwords and materials playing a key role in the development of the red category. The modern concept of a “red” category was the result of fusing the various reds into a single group.

Green, Green-Blue, and Green-Yellow

The Egyptian (*wādj*) and Akkadian (*warqu*) color terms that denote what today is referred to as “green” in English are derived from the same linguistic root but do not invariably denote the same color: the Sumerian *sig* is equivalent to Akkadian *warqu* and both also designate “yellow” (as in vegetation having both shades, but not in the sense of a green-yellow); the focus of Egyptian *wādj* is definitely in the “green” range (and not in

the green-blue range as originally suspected, cf. Schenkel in [1, 2, 12]).

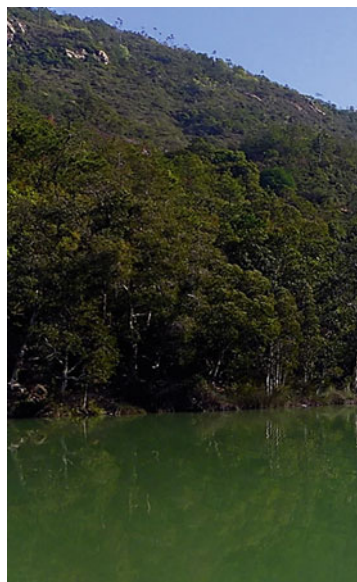
There was no generic word for what is today understood as “green” in Greek, although *xlōros* eventually came to mean something like “green”; however, the earliest use of color terms in Greek was not specific – or rather green was divided among a number of words, and not dominated by a “green-blue.”

qīng (青) – a Chinese term meaning “blue,” “green,” and “black,” but neither “dark” nor “green-blue” (with, however, a homonym including an additional radical [*qīng* 清] meaning “clear”!) – has not yet been found in the earliest Chinese inscriptions. Since the first millennium BC, *qīng* is used for “green,” “blue,” and “black”; only slightly later the word *lǜ*, today’s “green,” also appeared, so that in Chinese to some extent “green” has since been divided into “light” and “dark” green [5]. There is a still current old expression using the two different words for “green”: *qīng shān, lǜ shuǐ*; when translated narrowly, it would be “green mountains and green waters,” but actually means “(dark) green hills and (light) green lakes.” This phrase confirms an early and prolonged verbal division of the green part of the spectrum (Fig. 1).

Green clearly was not linguistically demarcated in a clear fashion with a unique term in any of the earliest languages except Egyptian – where its focus was in green as the spectrum is customarily divided, and not fused with other hues (Schenkel in [2, 12]). Otherwise, words for “green” designated not a vague color range, but rather distinct hues, i.e., in Mesopotamia “yellow” and “green,” in China “blue” and “green” – and **not** “green-blue” **nor** “green-yellow”; **nor** a general “dark.”

Yellow

Early evidence of terms glossing “yellow” is not common. In Egyptian, linguistic use of the word for “gold” (*nb.w*) for “yellow” began in the early third millennium BC, but remained uncommon; in Egyptian painting, however, the color yellow was frequently used to depict what was intended to represent gold where required and also the sun on occasion [8]. To what extent the Egyptian

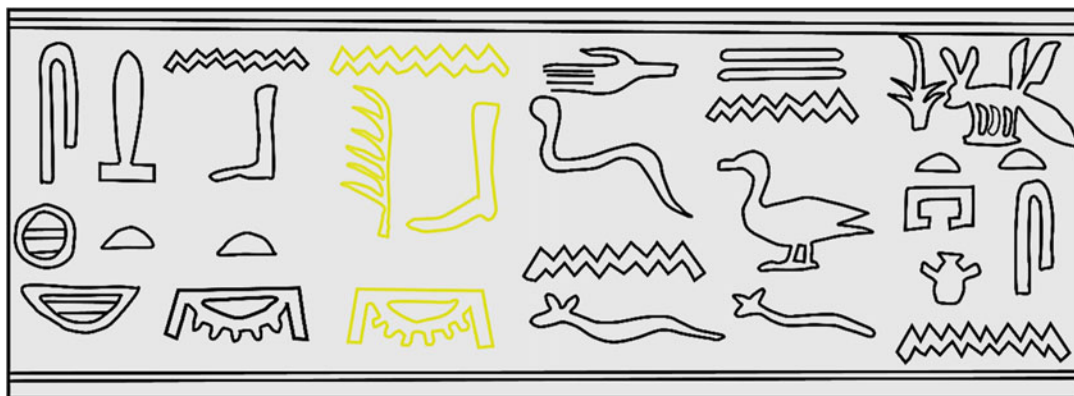


Ancient Color Terminology, Fig. 1 An image of southern China illustrating linguistic usage: 青山水色的天空 *qīng shān lǜ shuǐ lán sè de tiānkōng*; “dark green mountains, light green waters, the blue color of the heavenly vault.” *qīng* can also mean “blue” – as *ao*, the equivalent does in Japan today. Chinese has a highly differentiated vocabulary for green and blue, and the nuances reflect origins in the Bronze Age, but the abstraction and the distinction of hues is that of the era since the first millennium BC

“red”, *dšr*, was consciously used for “yellow” is not clear (Schenkel in [12]). The usage of gold in Akkadian texts with the meaning of “yellow” is even rarer than in Egyptian, and when used took the form of a simile; since the third millennium, Akkadian *duxšu* might have designated “yellow” taking the hue from calcite (Thavapalan in [12, 13]) (Fig. 2).

Although documented relatively early on in Chinese, it is unclear when *huáng* became a true color word meaning “yellow” (including “brown”); in antiquity, it was mixed with homonyms related to “royalty” and “jade” and thereby to “bright,” “brilliant,” as well as the color “yellow.” Chinese *jīn* (literally “gold”) seems to have been used for “yellow” from early on ([5, 18]), but not exclusively.

The later Greek words based on *xrusós*, “gold,” were frequently used to designate a color which was most probably “yellow.” Significantly,



Ancient Color Terminology, Fig. 2 Probably one of the oldest sentences formulated in writing in Ancient Egypt (at least decades before ca. 2600 BC), on a seal impression with the sun-god designated as *nbwj*, the “Golden One” (highlighted in the copy for readers: N.B. the original

impression was just mud; although the seal itself could have been made of lapis lazuli, there was no color in the impression, where the characters simply appeared as low relief in the mud), an early case of the use of “gold” for a yellow as a color term [17]

Mycenaean Greek Linear B and Classical Greek have two different words for a “yellow,” *ka-sa-to/ksanthós* and *ko-so-u-to/ksoûthos*, but neither is a real “yellow” [4], as the latter is presumably a yellow with a tinge of orange or red (as in saffron), and there is reason to suspect that *ko-so-u-to* (used to describe the colors of animal hides in Mycenaean Greek) might be closer to a “golden yellow” [16] or “brown,” depending upon context.

Blue

Blue is relatively well represented in most ancient languages – but not easily separable from materials (whether minerals, glass, or dyed textiles). In Akkadian (*uqnû*) and Egyptian (*xsbdj*), terms for lapis lazuli designated “dark blue” (both being loanwords related to the material, but derived from different words derived from some other languages). In Classical Greek, a term (*kuáneos*) for blue appearances is derived from the Mycenaean Greek Linear B *ku-wa-no*, a loan word apparently derived via Hittite *ku(wa)nnaš* and/or Ugaritic *iqnu* textually originally from the Akkadian *uqnû* (itself, however, a loanword of unknown origin). Akkadian used several other terms for blue, including *takiltu* (also known from Hebrew *tēkēlet*) for a more purple blue dye, and *xašmanum* for a “light blue” dye (from the Egyptian word for amethyst, *ḥsmn*, which was not used as a color word in Egyptian). In Egyptian,

turquoise (*mfkāt*) denoted “light blue.” Chinese *lán* is a term for “blue” color but appears quite late (in comparison to, e.g., Chinese “red,” “white,” “black,” “yellow”); *qīng* is also used for blue.

As a category, the modern English term “blue” evolved to ultimately eclipse the distinction established during early Antiquity (and still preserved in Russian) between light and dark blue.

Pink

Chinese is the only ancient language where one can be certain that *hóng*, a word that later came to mean “red” (the word used today) originally meant “pink” [18]. It is possible that Egyptian *dšr*, used here under “red” (as is the usual understanding), also included or meant “pink” (as it was always written with the flamingo ).

Table 1 offers an impression of what can be identified in the way of colors in these earliest languages.

Other Colors

Other colors are not clearly identified (or not hitherto recognized) in the earliest languages: the sweep of “yellow” could have included “brown”; the range of the “reds” will include “orange,” and “white” will include “gray.” Akkadian *makrû* – usually understood as a “red” – might mean “orange” on occasion. Probably

Ancient Color Terminology, Table 1 Summary of identified color terms from the earliest known languages

Language → Time →	Proto-Cuneiform 4th millennium BC	Sumerian 3rd-second millennium BC	Akkadian 3rd-first millennium BC	Egyptian 3rd-first millennium BC	Egyptian 1st millennium AD	Chinese End second millennium BC	Chinese End first millennium BC	Greek Late second millennium BC	Greek Late first millennium BC	Maya 1st millennium AD	Aztec Early second millennium AD
Color terms↓											
“White”	bar, ? u ₄	babbar	pešû	ḥḍj	oubaš	bái	bái	re-u-ko	leukós	*saq/sak	sak
“Black”	?gi	mí, gíg, gí ₆	šallamu	km	kame	xián	hēi	ma-ra	mélas	box, ehk’	ihk’
“Bright red”	si/u ₄ , ne ₆	sa ₃ , si/u ₄	sānu	dšr	trošrš	chì	chì	po-ni-ko-ro	eruthrós	*kaq/chak	tlapalli
“Green”	sig ₇	sig ₇	warqu	wādj	ouot		qīng		xlōrós		quiltic
“Yellow”	?gi	sig ₁₇	?duššu	nb.w		huáng	huáng	ka-sa-to	ksanthós, ksanthos	*q’an/k’an	cozic
“Dark blue”		su ₆ -za-gin-na	uqniātu	xšbdj				ko-so-u-to	kuáneos		texotic
“Grue,” “blue-green”										ya’aš, ya’ax	xoxotic
“Light green”									prásinos		
Other											
“blues”			xašmānu, tukiltu	jrtjw, mfkāt			qīng	ku-wa-ro	glaukós	yax	xox, ocic, mallaltic
Other “reds”		?xuš; xuš-a	pehum, ruššu, xuššu	rwdj, mss, tjms, mroš	mroš			er-ru-to-ro	rhodeis		chichiilitic,
“Pink”		ša ₄ , suo, sa ₅				hóng		ka-ra-u-ko			tlaztalehualli
“Orange”			?makrû								xochipaltic
“White,” “light,” “bright,” “dazzling”	(babbar?)	ara, bar, ḥáda, dalla, kára, kug, pirigrín, še-er, tán, zalag	pešû	ḥḍj, tjint		ming, qīng	ming, qīng				
“Dark”	(gege?)	dara, ge, gíg/gege, kúkku, mi, šuš/šú, mes	šallamu, tarku	kk.w			an, huì, méng, míng			ek	

influenced by materials, shades distinguished within green, red, yellow, and blue were regarded as distinct colors, before the addition of additional colors.

Theory of Ancient Color Term Emergence

Debate regarding the sequence of the emergence of color in early color terminology is constrained by color term emergence theories (see also Encyclopedia of Color Science and Technology entry on ► [“Berlin and Kay Theory”](#)) using contemporary and recent language data to project modern color term emergence backward into history, rather than following the path forward, using the relevant ancient data discussed here.

Of extreme importance is that in ancient Egypt, the sun is occasionally called “red” and painted as such. This does not imply a “red-yellow,” but rather the use of a single term to designate what in modern languages are treated as different colors. The same goes for the Mesopotamian *warqu* “green” used for “yellow” and “green” as well as the Chinese *qīng* meaning “blue” and “green.” The reasons behind this do not lie in perception so much as expression.

That the Chinese had a “pink” before they had a red is but the tip of the iceberg (since “pink” should come much later according to Berlin & Kay). Most important is that “bright”/“white” and “dark”/“black” were not necessarily distinguished as color terms before the emergence of a “red.” Equally important is that before adding words for “brown,” “orange,” etc., the ancient languages divided “blue,” “red,” and “yellow,” with several different terms – with particularly disconcerting results for the value of “green.”

In later Western languages, “black,” “white,” “red,” “green,” “yellow,” and “blue” were each largely consolidated under one single word – as Berlin & Kay stipulate and in contrast to the earlier system where early languages had multiple “reds,” as exemplified in Schenkel’s work indisputably confirming the two or three reds in Egyptian before the acquisition of other colors.

Furthermore, there is less coherence when arguing abstraction, as this is less evident in the earliest languages where materials play a major role (see also Encyclopedia of Color Science and Technology entries on ► [“Materials and the Origins of Abstract Color Naming”](#); ► [“Evolution of Color Expression”](#)). This development of abstraction seems to have begun in parallel in both Chinese and Greek in the middle of the first millennium BC – but in neither of the ancient forms of these languages was any kind of systematic vocabulary developed that corresponds to the modern understanding of “green” or “blue.” Although – exceptionally – in both Chinese and Greek, “black,” “white,” and “red” seem to have been largely accommodated with suitable salient terms by the end of the first millennium BC, really adequate salient words for “yellow,” “blue,” and “green” did not emerge so clearly.

Warburton [7] argues that the earliest acquisition sequence was not an issue related to individual specific languages, but rather a shared phenomenon based on shared terminology, concepts, and materials. In this scheme, “ocher” (as dark red) will have been the earliest color used (in the Paleolithic, hundreds of thousands of years ago), followed by “greenstones” (in the earliest Neolithic, from 10,000 years ago), and “gold,” “silver,” “lapis lazuli,” “carnelian,” and “obsidian” (in the following late Neolithic and earliest Bronze Age) and then by turquoise (before 4500 years ago), amethyst, and other stones. The imitations in glass and faience bore the names of the stones, but they will hardly have consistently reproduced the hues recognized in languages today; dying textiles will have contributed yet more hues associated with specific terms – not all of which designated what would today be viewed as comparable colors. Initially, this led to a variety of color terms. “Black” and “white” may not have emerged from words for “dark” and “bright” before “red,” but rather linguistically crystallized as individual words after the understanding of “red,” probably the first linguistically distinct independent color recognized. Colors such as “pink” (in Chinese) and “orange” (in Akkadian, Egyptian, and Mycenaean) may have been recognized and designated with



Ancient Color Terminology, Fig. 3 Beads of gold, amethyst, carnelian, faience, and glass imitation of amethyst found in Crete at the end of the Minoan period (mid-second millennium BC [19]). Precious materials

specific names, but what they specified may be foreign to modern observers (Fig. 3).

Summary

Two of the three earliest linguistically recognized colors – “red,” “black,” “white” – may have been derived from materials of the ordinary existing world (perhaps ocher and soot), but the link between “white,” “silver,” and “purity” means that in linguistic terms, the third color already took us into the domain of precious materials. Henceforth, “precious materials” played an important role in the development of the advanced abstract color terminology (with stones of green, yellow, and blue along with gold), but with carnelian “red,” a precious material was intimately associated with a designation for a the shade of red already recognized earlier (ocher), splitting “red” and linking the perception and expression of color to hard, bright objects with intense clear hues all bound to materials and loanwords accompanying objects before, during, and after the period of the earliest recorded languages. The earliest recorded languages – Proto-Cuneiform, Sumerian, Egyptian, and Akkadian – reveal virtually no purely abstract colors with a focus on a recognizable color, but the art reveals a search for both pure reds, whites, and blacks while linguistically and artistically distinguishing different reds, blues, and greens. Abstract words appear

were traded and the names of faience and glass imitations adopted from the names of real precious stones, and used to designate textile dyes, bringing loanwords associated with strong hues into circulation

gradually in Chinese and Greek in the first millennium BC. Warburton suggests that the materials dominated and that the concept of abstract color terminology never developed in the Ancient Near East; yet the flow of loanwords meant that abstract terminology may have already been embryonic in the Aegean by ca. 1400 BC [8].

There is a clear difference between the descriptive uses of color in literature and economic texts. Although the sources are not balanced, there is little evidence of color from the earliest administrative texts in proto-cuneiform whereas color terminology appears in the administrative texts written from the third millennium onward.

Loanwords and shared etymologies play important roles in the emergence of ancient color terms in that many of the earliest words are related to designations of materials. Terms for “white” and “black” do not stand at the origins, and the red, blue, and green ranges were partitioned before being lumped together. Only later did the further color words begin to crystallize (in the last two millennia). The concept of “blue-green” is not documented in the earliest languages and does not really evoke any consistent color significance among potentially relevant living sources [5] suggesting that the concept may never have existed.

Since the third millennium BC, there has clearly been a gradual increase in understanding colors as hues and associating specific terms with specific parts of the spectrum. Even the simplest

color vocabulary is a very recent cognitive development: the use of complex abstract color language can hardly be extended back 5000 years [14], being instead an historical development in relatively recent millennia, leading to the emergence of abstraction in recent historical times. The Greeks acquired color concepts and terminology from the Near East through the diffusion of materials, ideas, and terminology. Their vague abstract ideas influenced the color lexicons of more recent languages.

Cross-References

- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Cross-cultural Communication](#)
- [Effects of Color Terms on Color Perception and Cognition](#)
- [Evolution of Color Expression](#)
- [Mesoamerican Color Survey Digital Archive](#)
- [World Color Survey](#)

References

1. Schenkel, W.: Die Farben in Ägyptischer Kunst und Sprache. *Z. Ägyptische Sprache*. **88**(131–147) (1963)
2. Schenkel, W.: Color terms in ancient Egyptian and Coptic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 211–228. Benjamins, Amsterdam (2007)
3. Bulakh, M.: Basic color terms from proto-semitic to old ethiopic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 247–261. Benjamins, Amsterdam (2007)
4. Blakolmer, F.: Zum Charakter der frühägäischen Farben: Linear B und Homer. In: Blakolmer, F. (ed.) *Österreichische Forschungen zur Ägäischen Bronzezeit 1998: Akten der Tagung am Institut für Klassische Archäologie der Universität Wien 2.–3. Mai 1998*, pp. 225–239. Wiener Forschungen zur Archäologie, Vienna (2000)
5. Xu, W.: *A Study of Chinese Colour Terminology*. Lincom, Munich (2007)
6. Baines, J.: Color terminology and color classification: ancient Egyptian color terminology and polychromy. *Am. Anthropol.* **87**, 240–262 (1985)
7. Warburton, D.A.: The theoretical implications of ancient Egyptian colour vocabulary for anthropological and cognitive theory. *Lingua Aegyptia*. **16**, 213–259 (2008)
8. Warburton, D.A.: Colors in bronze age Egyptian art and language. In: Brinkmann, V., Primavesi, O., Hollein, M. (eds.) *Circumlitio: the Polychromy of Antique and Mediaeval Sculpture*, pp. 170–187. Liebighaus, Frankfurt (2010)
9. Berlin, B., Kay, P.: *Basic Color Terms: their Universality and Evolution*. University of California Press, Berkeley (1969)
10. Ierodiakonou, K.: Plato's theory of colours in the *Timaeus*. *Rhizai*. **2**, 219–233 (2005)
11. Goldman, R. (ed.): *Essays in Global Color History: Interpreting the Ancient Spectrum*. Gorgias Press, Piscataway (2016)
12. Thavapalan, S., Warburton, D.A. (eds.): *The Value of Color: Material and Economic Aspects in the Ancient World*. *Berlin Studies of the Ancient World* 70, Berlin (2019). <https://doi.org/10.17171/3-70>
13. Thavapalan, S.: *The Meaning of Color in Ancient Mesopotamia. Culture & History of the Ancient near East*. Brill, Leiden (2019)
14. Warburton, D.A.: Colourful meaning: terminology, abstraction, and the near eastern Bronze Age. In: Johannsen, N., Jessen, M., Jensen, H.J. (eds.) *Excavating the Mind: Cross-Sections through Culture, Cognition and Materiality*, pp. 183–208. Aarhus University Press, Aarhus (2012)
15. Houston, S.D., Brittenham, C., Mesick, C., Tokovinine, A., Wariner, C.: *Veiled Brightness: a History of Ancient Maya Color*. University of Texas Press, Austin (2009)
16. Liddel, H.G., Scott, R., Jones, H.S.: *A Greek-English Lexicon*. Oxford University Press, Oxford (1958)
17. Kahl, J.: “Ra Is my Lord”: Searching for the Rise of the Sun God at the Dawn of Egyptian History. Harrassowitz, Wiesbaden (2007)
18. Schuessler, A.: *ABC Etymological Dictionary of Old Chinese*. University of Hawai'i Press, Honolulu (2007)
19. Evans, A.: *The Palace of Minos at Knossos. Volume IV, Part II*. Macmillan, London (1935)
20. Englund, R.K., Nissen, H.J.: *Die Lexikalischen Listen der Archaischen Texte aus Uruk*. Gebr Mann, Berlin (1993)
21. Landsberger, B.: Über Farben im Sumerisch-Akkadischen. *J. Cuneif. Stud.* **21**, 139–173 (1967)

Apparent Magnitude – Visual Magnitude

- [Apparent Magnitude, Astronomy](#)

Apparent Magnitude, Astronomy

Ken Tapping
D.R.A.O, National Research Council Canada,
Penticton, BC, Canada

Synonyms

Apparent magnitude – visual magnitude; Opacity – extinction

Definition

The apparent magnitude of a celestial object is a measure of how it looks to the eye. However, to make such a measurement useful to other observers at other locations and other times a more rigorous and transferrable standard is needed. This system needs to accommodate the different colors of celestial objects and the widely varying color sensitivity of the human eye. To achieve this, standardized systems of color filters have been established, which used together with standardized photometers, yield a useful system which not only defines the brightnesses of celestial objects in a consistent way, but also forms a useful basis for the spectral classification of stars.

Glossary of Terms

- Elevation: Angle between the geometrical horizon (as indicated by a spirit level) and the direction of the star or other object in question.
- Extinction: The degree to which starlight is weakened by the medium through which the light travels from the star to the observer. It is similar in meaning to “opacity” but is usually expressed in magnitudes. Zero means no extinction; otherwise the extinction indicates how many magnitudes fainter the object looks due to opacity of the interstellar medium and the atmosphere.

- Opacity: The degree to which light (or other optical radiation) is blocked when passing through the material in question. It varies between zero for something that is completely transparent and infinity for something completely opaque.
- Orthochromatic: This term refers to a type of photographic emulsion that was widely used in the earlier days of photography. It was very sensitive to blue and green light but insensitive to red light.
- Photometry: The measurement of quantities referred to light (visible optical radiation) as evaluated according to a given spectral luminous efficiency function (of the human eye). In astronomy, it applies to the measurement of the brightness of astronomical objects.
- Photosphere: The layer in a star’s atmosphere from which almost all of the heat and light are radiated. Because it lies at a fairly abrupt change in optical thickness, it looks almost like a solid surface and is often referred to colloquially as the “star’s surface.”
- Transmissivity: The fraction of the input energy passing through an attenuating device or layer. It varies from unity, for a completely transparent medium, to zero, for something completely opaque.
- Zenith angle: Angle between the local zenith and the direction in question.

Background

At the most fundamental level, the apparent magnitude of an object is a measure of how bright it looks. This is a function of the properties of the light source, its distance, the nature of the medium through which the light travels to the observer, and inevitably the properties of the observer or device used to make the measurement. When minimizing the subjective effects and trying to make the measurement more descriptive of the source and less related to the local circumstances of the observer, the parameter in question inevitably becomes less “apparent” but more useful. The history of “apparent magnitude” reflects such an evolution.

At a qualitative level, at a particular time and location, under the atmospheric conditions existing at that time, a specific observer can assess that some objects in the sky look brighter than others and possibly estimate by how much. This information may be useful to other observers present at the same location at the same time. However, recording such information for later consideration is difficult, and passing it in any usable form to another observer at another location for comparison of observations is even more so.

The development of the concept of “apparent magnitude” marks the beginning of an evolution toward producing quantities that are more intrinsic to the object being observed and can be combined with other observations. In the process, the term “apparent” becomes more a designator of a particular type of data rather than something literally true.

The flux from a distant star arriving at the top of the Earth’s atmosphere is given by

$$F_{00}(\lambda) = \alpha(\lambda) \frac{W^*(\lambda)}{4\pi R^2} \quad (1)$$

where W^* is the energy output of the star, R its distance from us, and α is the transmissivity of the interstellar medium along the path from the star to Earth. Note the wavelength dependence of these parameters. In this discussion, the “00” suffix identifies a quantity measured at the top of the Earth’s atmosphere.

Because the same value of F_{00} would be incidentally equal upon all points at the top of the Earth’s atmosphere, it would be common to all terrestrial observers. Therefore, in these discussions, F_{00} can be considered as the starting point.

In the development of the apparent magnitude concept, a number of issues arise. Firstly, since this is a comparative system, a defined calibration datum is required. In addition, starlight is not monochromatic; the atmosphere’s transmissivity (or opacity) varies with wavelength and so does the sensitivity of the detectors, sensors, or measurement devices used.

Quantifying Apparent Brightness

The concept of assigning “magnitudes” to stars was first proposed by the ancient Greeks, and has been attributed to Hipparchos [1], who was a Greek astronomer, geographer, and mathematician who lived about 190–120 BC. Accommodating the natural nonlinear response of our eyes to brightness, he assigned a magnitude scale where the brightest stars were categorized as being of the First Magnitude ($m = 1$), and the faintest stars visible to the unaided eye were defined as being of the Sixth Magnitude ($m = 6$), with each intervening magnitude corresponding to half the brightness of the next brighter class and twice the brightness of the next fainter. In this system, if the energy fluxes reaching from two stars of magnitude m_1 and m_2 are respectively F_1 and F_2 , the relationship between these quantities is

$$\frac{F_1}{F_2} = 2^{-(m_1 - m_2)} \quad (2)$$

This classification scheme for the apparent magnitude of stars was used for some time. Ptolemy used it in *The Almagest*, [2], his compendium of mathematical and astronomical information. Since the magnitudes as defined are exponents, taking logarithms of Eq. 2 puts it into a more convenient form

$$\begin{aligned} \log_{10} \left(\frac{F_1}{F_2} \right) &= -\log_{10}(2)(m_1 - m_2) \\ &= -(0.30102996 \dots) \\ &\quad \times (m_1 - m_2) \end{aligned} \quad (3)$$

so

$$(3.321928 \dots) \log_{10} \left(\frac{F_1}{F_2} \right) = -(m_1 - m_2) \quad (4)$$

The logarithm of the ratio of brightnesses is simply related to the magnitude difference. However having irrational numbers as the scaling factors is inconvenient, and in 1856, Norman Robert Pogson [3] redefined the magnitude scale in a more useful form. In Hipparchos’ system, a first

magnitude star is defined as being 64 times brighter than one of the sixth magnitude. Pogson changed the definition of the magnitude scale, defining a sixth magnitude star to be 100 times fainter than a first magnitude star. So the ratio of brightness between two adjacent magnitudes became

$$P = 100^{\frac{1}{5}} = 2.5118864315 \dots \quad (5)$$

This number has become known as *Pogson's Ratio* (P). Despite the apparent messiness of this number, it has an advantage in calculations. P is almost always used in the forms $\log_{10}(P)$ and $1/\log_{10}(P)$, which equal exactly 0.4 and 2.5, respectively. Thus, Eqs. 2 and 3 become

$$\frac{F_1}{F_2} = P^{-(m_1 - m_2)} \quad (6)$$

$$\begin{aligned} \therefore \log_{10}\left(\frac{F_1}{F_2}\right) &= -\log_{10}(P)(m_1 - m_2) \\ &= -0.4(m_1 - m_2) \end{aligned} \quad (7)$$

so

$$(m_1 - m_2) = -2.5 \log_{10}\left(\frac{F_1}{F_2}\right) \quad (8)$$

Since the magnitude system describes relative brightnesses, it needs to be based upon a calibration reference: a star to which a magnitude value has been assigned. Pogson first chose Polaris (also known as the Pole Star or North Star) as a calibrator, to which he assigned a magnitude of 2. However, when this star proved to be variable in brightness, he selected Vega, which he defined as having an apparent magnitude of 0. In practice, Vega is a better choice because it is available to observers over much of the world, and since it varies in elevation with time at any given site, other stars at similar elevations at that time can easily be set up as secondary or transfer calibration standards.

The nature of the magnitude scale makes it able to accommodate stars brighter than magnitude 0, which gives them negative magnitudes. For

example, Sirius, the brightest star in the constellation of Canis Major, and the brightest star in our sky other than the Sun, has a magnitude of -1.5 and the Sun a magnitude of about -26 . Fainter stars, with magnitudes greater than 6 can also be accommodated. Binoculars or telescopes will reach objects far fainter than the eye can discern, which therefore have magnitudes greater than 6.

Wavelength Issues

Stars are not sources of monochromatic light; they emit a broad spectrum extending across the nominal 400–800 nm band of visible wavelengths and beyond in both the longer and shorter wavelength directions. The apparent color of a star depends upon the wavelength at which the peak brightness occurs, which in turn depends upon its photospheric temperature. The consequent star colors range from red to blue (and into the infrared for very cool stars).

The interstellar and atmospheric transmissivities are wavelength dependent, and in general so is the sensitivity of devices typically used to measure light intensity. This variability is particularly marked in the case of the human eye, which may differ between individuals and also with the same individuals, depending upon health, fatigue, etc. A classic example of the need to take this wavelength dependence into account is evident in old photographs of the constellation of Orion, made using orthochromatic photographic emulsions, which have extremely low sensitivity to the red end of the visual spectrum. Betelgeuse (α Orionis), the brightest star in the constellation (a red giant), is not visible, whereas Rigel, a blue-white star, is very conspicuous. A more extreme, although nonastronomical example is in a photograph (again using an orthochromatic emulsion) of the cab of an early steam locomotive, looking into the open firebox, which shows no fire at all, although the boiler pressure gauge shows a vigorous fire must be present. To maintain the utility of the apparent magnitude observations, it is necessary to deal with wavelength dependencies, but this needs to be done in a manner that is

both standardized and does not overly complicate the measurement process.

Stellar Spectral Types

The concept of “stellar spectral type” is, to a large extent, an indicator of a star’s photospheric temperature, i.e., its color. This way to describe stars is therefore intricately mixed with the magnitude/color issue, so a short description is included here. Stars come in a wide range of colors, brightnesses, and masses and with different spectral line signatures, so it was deemed useful to develop some sort of classification framework toward providing a coherent overall picture of star types and stellar evolution. The first attempts were made at a time when stellar astrophysics was not a highly developed subject. One early system to classify stars used alphabetic labels. However, as knowledge improved, the trend was to classify on the basis of photospheric temperature and spectral line signatures. Some classes were found to be redundant, and the sequence of remaining classes had to be rearranged, so they were no longer in alphabetical order. The new system is in order of decreasing temperature and became O, B, A, F, G, K, M, L, T, Y. This arbitrary sequence of letters is remembered using the mnemonic “Oh Be A Fine Girl and Kiss Me. Like This? Yes!” This sequence is further subdivided using numerical interpolation (e.g., G2, A5, M0, etc.). The classes and their corresponding photospheric temperatures and colors are summarized in Table 1. In the case of the Y0 class at 600 K, these cool stars are usually described as “brown,” although the color would be more accurately described as an extremely dull red or perhaps not visible at all at optical wavelengths since its spectral peak lies in the infrared region.

The UBV(RI) Magnitude System

Stellar spectra are generally (almost) blackbody spectra corresponding to the stars’ photospheric temperatures, with superimposed absorption

Apparent **Magnitude,** **Astronomy,**
Table 1 Photospheric temperatures/colors and spectral types

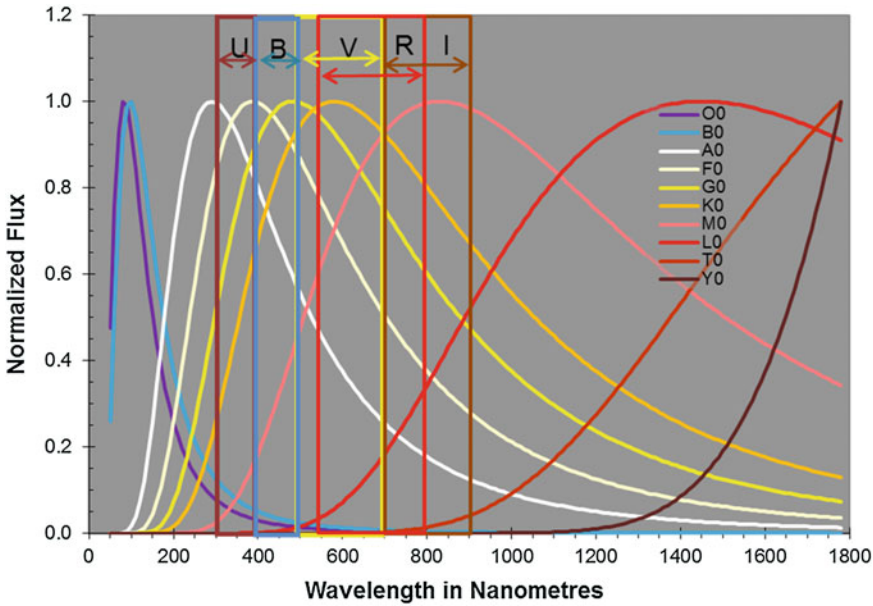
Photospheric temperature (K)	Spectral type	Color
35,000	O0	Blue
30,000	B0	Blue white
10,000	A0	White
7,500	F0	Pale yellow
6,000	G0	Yellow
5,000	K0	Orange
3,500	M0	Red
2,000	L0	Dull red
1,200	T0	Very dull red
600	Y0	“Brown”

lines. Magnitude determination applies to the thermal emission only. Figure 1 shows blackbody spectra for the photospheric temperatures and spectral types listed in Table 1.

The required amplitude and wavelength-related information is included in a star’s spectrum. Stellar spectra are the primary data source for most stellar astrophysics. However, each spectrum contains a lot of information, and when studies involving large numbers of stars, or where measurements that are relatively simple and reproducible are needed, an alternative system would be useful.

To meet these needs, Johnson and Morgan [4] developed the three-filter, UBV system, which is still widely used. This method uses magnitude determinations made using three standard filters, designated U (ultraviolet), B (blue), and V (visible), respectively. To better describe cool, red stars, two additional bands have been added: R (red) and I (infrared). The objective is to describe the thermal spectrum using a small list of standardized parameters. The nominal wavelength ranges (bandwidths) are listed in Table 2.

The shapes of the filter passbands for devices currently in use are far from rectangular, but there has been an effort to at least standardize the filters to improve comparability of data between sources. The I and R bands are not so cleanly



Apparent Magnitude, Astronomy, Fig. 1 Blackbody spectra corresponding to the photospheric temperatures of stars of spectral types O0, B0, A0, F0, G0, K0, M0, L0, T0, and Y0. The nominal bandwidths of the U, B, V, R, and I filters are shown. The spectra are all normalized

Apparent Magnitude, Astronomy, Table 2 Nominal filter bandwidths for stellar magnitude measurements

Band designation	Band name	Low end (nm)	High end (nm)	Bandwidth (nm)
U	Ultraviolet	320	400	80
B	Blue	400	500	100
V	Visible	500	700	200
R	Red	550	800	250
I	Infrared	700	900	200

delineated. The I band overlaps the V band, and the R band overlaps the I band. Some institutions have added additional filter bands to suit their particular research needs.

Table 1 also shows the wavelength ranges passed by the U, V, B, R, and I filters. These rectangular passbands are not representative of the passbands of existing filters. The bandwidths are also shown using arrows, so the overlapping bands are more clearly discernible.

For most purposes, the U, B, and V magnitudes are sufficient, and it is possible to describe the photospheric thermal spectrum of most stars using V and the magnitude differences U–B and B–V. Since the magnitudes are logarithmic

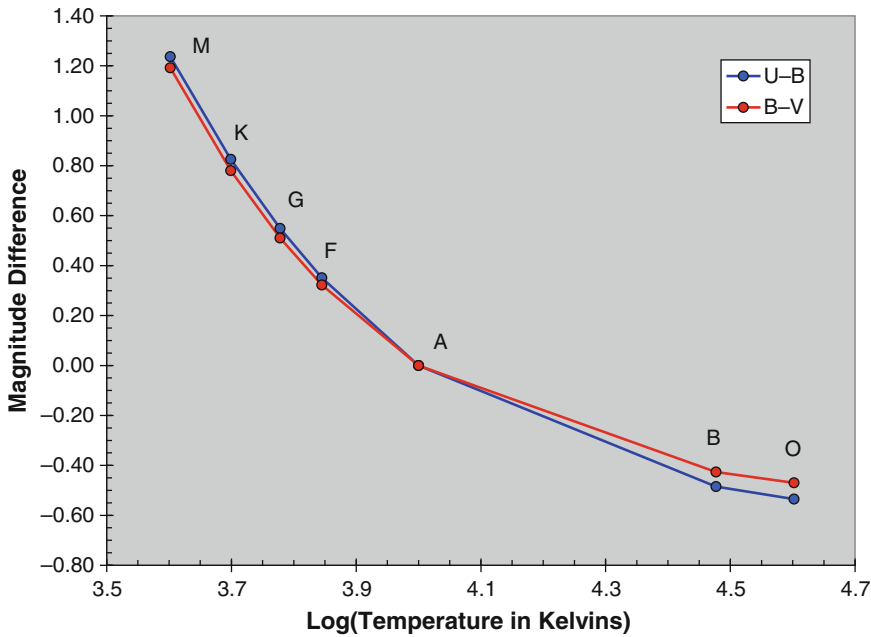
quantities, the differences actually describe brightness ratios, i.e.,

$$(U - B) = -2.5 \log_{10} \left(\frac{\frac{P_U}{\Delta\lambda_U}}{\frac{P_B}{\Delta\lambda_B}} \right) \quad (9)$$

and

$$(B - V) = -2.5 \log_{10} \left(\frac{\frac{P_B}{\Delta\lambda_B}}{\frac{P_V}{\Delta\lambda_V}} \right) \quad (10)$$

where P is the total power passing through the filter. Conventionally, the values are offset so that a star of spectral type A0 is characterized by U–B



Apparent Magnitude, Astronomy, Fig. 2 U-B and B-V plotted against the logarithm of the photospheric temperature for the blackbody spectra in Fig. 1. The spectral

types of the stars modeled are shown. In each case the types are M0, K0, etc.

and $B-V = 0$. If the filter passbands are assumed to be rectangular, the spectra in Fig. 1 produce the U-B and B-V values shown in Fig. 2. It can be seen that these values provide a good estimator for the star's photospheric temperature. The letters above the data points on the graph are the spectral types of stars with those particular photospheric temperatures. Since stellar spectral types are temperature dependent, the UVB photometric system provides an estimator for spectral type.

However, in reality, filter passbands are not rectangular, and stellar flux densities may vary dramatically with wavelength, so simply averaging over assumed rectangular passbands will not produce useful results. Unfortunately, making filters with rectangular passbands is very difficult, and the more effort dedicated to achieving this also makes the filters harder to reproduce, which is a critical issue where standardization between instruments and observatories is important.

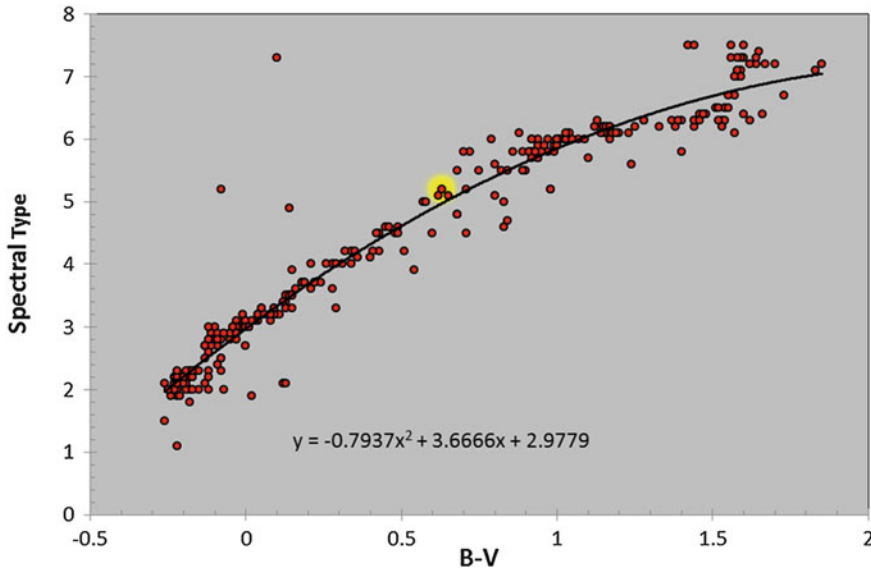
For this reason, a set of standardized filters are used. These are generally those used by Johnson and Morgan [4]. They do not have rectangular passbands but have the great advantage of being

fairly easy to reproduce. The wide adoption of these standard filters for stellar photometry makes it possible to combine data from different sources.

B-V and Stellar Spectral Types

Figure 3 shows a plot of spectral type against B-V for the stars listed in the bright star section of the *Observer's Handbook* of the Royal Astronomical Society of Canada (RASC) [5]. For convenience in plotting, the spectral types O0, B0, A0, F0, G0, K0, and M0 are represented by the numbers, 1, 2, 3, 4, 5, 6, and 7 respectively. The subdivisions within each spectral class are coded as a decimal; for example, stars of spectral classes O5, A0, F6, G2, K5, and M3 would respectively be represented as 1.2, 3.0, 4.6, 5.2, 6.5, and 7.3. The point surrounded by the yellow halo is the Sun. The empirical equation can be used to estimate the spectral type from B-V.

Today, the term "apparent magnitude" is generally taken to be equivalent to "visual



Apparent Magnitude, Astronomy, Fig. 3 Numerical representation of spectral types plotted against B–V values for the stars in the RASC list of brightest stars

magnitude”: a magnitude measurement made using a V filter.

Absorption and Scattering

To reach terrestrial observers, the starlight has to travel through the interstellar medium and then through the Earth’s atmosphere. Both of these media contain dust, molecules, and atoms that absorb or scatter some of the starlight, with the extent of these processes usually being wavelength dependent. The result is generally a weakening and reddening of the light. There are three main processes:

- Scattering at wavelengths coinciding with molecular or atomic resonances, where the directional light drives electrons to higher energy states and then is reradiated isotropically.
- Rayleigh scattering, which is scattering by particles smaller than the wavelength of the light (mainly atoms and molecules). This is strongest at shorter wavelengths and is what causes

the sky to look blue and sunsets to be rich in reds and golds.

- Scattering off larger particles: This shows much less of a variation with wavelength. In the atmosphere, the main scattering agents are aerosols, composed of small droplets of water or pollutants. This mode of scattering is less likely in interstellar space but may occur where there are clouds of larger particles. Fog and mist are good examples.

If an intrinsically stellar measurement is needed, then both the interstellar and atmospheric effects on the starlight have to be determined and removed. Although even the densest interstellar clouds are more rarefied than the best achievable laboratory vacuum, the very large distances the light has to traverse means it has to pass through an enormous amount of material. However, in this entry the emphasis is on making measurements that are comparable between observers and observing sites, so the measurements need only be referred to the top of the Earth’s atmosphere, so interstellar effects are common to all observers and not further discussed here.

The attenuation imposed by the atmosphere is by processes similar to those attenuating and scattering the light in the interstellar medium. However, an important difference is that in the case of the interstellar medium, the light path from star to the Earth is common to all observers and changes very slowly with time, whereas in the case of the atmosphere, the rotation of the Earth changes the light path through the atmosphere quickly enough to offer a means to estimate the stellar magnitude that would be observed in the absence of atmospheric attenuation. It is assumed that the transmissivity of the atmosphere is related to the total mass of atmosphere that the light passes through. If we look in the direction of the zenith (zenith angle = 0), we are looking through a certain total mass of air:

$$M_0 = \int_0^\infty \rho(h)dh \quad (11)$$

where $\rho(h)$ is the atmospheric density as a function of height. The “0” suffix indicates a quantity measured when looking at the zenith. The *transmissivity* of the atmosphere (the fraction of light arriving at the top of the atmosphere that reaches the ground) from an object at the zenith is defined as

$$F_0(\lambda) = F_{00}(\lambda) \exp(-\eta(\lambda)M_0) \quad (12)$$

where η is an absorption parameter and M_0 the mass of air through which the light is passing. The flux density from the star is a function of wavelength and so is the atmospheric attenuation.

Assuming the magnitude measurements are restricted to zenith angles smaller than about 65° (elevation angles larger than 25°), the atmosphere can be treated as a stratified slab. In this case, the air mass through which the light passes is a simple function of zenith angle:

$$M_z = M_0 \sec(z) \quad (13)$$

It is unlikely that there will be any need to separate M_0 and η , so they can be combined into a single parameter $\beta(\lambda)$, and so for a zenith angle z

$$F_z(\lambda) = F_{00}(\lambda) \exp(-\beta(\lambda) \sec(z)) \quad (14)$$

We have two ways to define the effect of the atmosphere on the light. The *transmissivity* (α) is the fraction of the incident light reaching the ground:

$$\alpha = \frac{F_z(\lambda)}{F_{00}(\lambda)} \quad (15)$$

An alternative is define an *opacity* (ξ), where

$$\xi(\lambda) = \beta(\lambda) \sec(z) \quad (16)$$

So

$$\begin{aligned} \alpha &= \exp(-\xi) \\ \text{or} \\ \xi &= -\log(\alpha) \end{aligned} \quad (17)$$

As the atmosphere ranges from completely transparent to completely opaque, the transmissivity (α) ranges from one to zero and the opacity (ξ) goes from zero to infinity.

The slab atmosphere model offers a means to estimate the magnitude of the star in the absence of the atmosphere, using a method attributed to Pierre Bouguer and discussed by Dufay [6]. Using Eqs. 7 and 12,

$$m_z(\lambda) - m_{00}(\lambda) = -2.5 \log_{10} \left(\frac{F_{00} \exp(-\beta(\lambda) \sec(z))}{F_{00}} \right) \quad (18)$$

So

$$\begin{aligned} m_z(\lambda) &= m_{00}(\lambda) \\ &+ 2.5 \log_{10}(e) \beta(\lambda) \sec(z) \end{aligned} \quad (19)$$

where m_z is the measured stellar magnitude at a zenith angle z and m_{00} is the magnitude of the star before it is attenuated by the atmosphere. Equation 16 shows a simple, linear relationship between m_z and $\sec(z)$. Therefore, if a series of magnitude measurements for a star are made over a range of zenith angles, a plot of m_z against $\sec(z)$

will have a slope $2.5 \log_{10}(e) \beta(\lambda)$ and intercept $m_{\infty}(\lambda)$, the magnitude of the star as it would be measured at the top of the Earth's atmosphere. This process is repeated for the U, B, and V filter bands, to obtain the U, B, and V magnitudes corrected for atmospheric absorption.

Summary

“Apparent magnitude” is literally a measure of how bright something looks. As such it includes the effects of observing conditions, locale, and the nature of the observer. In addition, stars come in a range of colors, depending on their temperatures. To make this measure useful to all terrestrial observers, it has to be less apparent and more objective. This is achieved by making stellar magnitude measurements using three filters, each covering a standard wavelength range, designated U, B, and V. The latter, the visual magnitude, covering the bandwidth 500–700 nm, replaces apparent magnitude.

Cross-References

- [Color Magnitude Diagrams](#)
- [Color Pollution](#)
- [Galaxy Color Magnitude Diagram](#)
- [Hertzsprung-Russell Diagram](#)
- [Psychological Color Effects](#)
- [Umov Effect](#)

References

1. Taub, L.C.: Hipparchus. In: John Lankford, J. (ed.) *A History of Astronomy: an Encyclopedia*. Garland, New York (1997)
2. Pederson, O.: *A Survey of the Almagest*. Springer. ISBN: 978-0-387-84825-9 (2011)
3. Pogson, N.R.: Magnitudes of 36 of the minor planets for the first day of each month of the year 1857. *MNRS*. **17**, 12 (1856)
4. Johnson, H.L., Morgan, W.W.: Fundamental stellar photometry for standards of spectral type on the revised system of the Yerkes spectral atlas. *Astrophys. J.* **117**, 313 (1953)
5. Garrison, R.F., Karmo, T.: The brightest stars. In: Chapman, D. (ed.) *Observer's Handbook of the Royal Astronomical Society of Canada*, RASC Publications, Toronto, pp. 273–283 (2013)
6. Dufay, J.: *Introduction to Astrophysics*. The Stars Dover Press, Chapters 1, 2 and 3. Dover Press, New York (1964)

Appearance

José Luis Caivano¹ and Paul Green-Armytage²

¹Secretaría de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

²The School of Media, Creative Arts and Social Enquiry, Curtin University, Perth, WA, Australia

Synonyms

[Color appearance](#); [Perceptual appearance](#); [Total appearance](#); [Visual appearance](#)

Definition

Visual appearance refers to how things look to observers. Objects have physical properties, and in this regard, optical properties are usually most relevant together with size, texture, shape, etc. But what finally matters for vision is the appearance that objects have for an observer (human observers will be mainly referred to in this article; however, this is not restrictive to human beings since there are many other living species that have similar visual processing systems). The American Society for Testing and Materials (ASTM) as well as the International Commission on Illumination (CIE) define appearance as the aspect of visual experience by which objects are recognized [1, 2]. The ASTM specifies, furthermore, the definition of appearance from a psychophysical point of view: “perception in which the spectral and geometric aspects of a visual stimulus are integrated with its illuminating and viewing environment.” The physical properties have, of course, influence on the way things look, but also the

neurophysiology and psychology of vision play a very important role in determining how objects and the world appear to humans. Thus, it is important to understand visual appearance not merely in regard to physical aspects but as a sensation produced by the visual system when processing stimuli channeled by means of light – visible radiation. Visual appearance, understood as a visual sensation, i.e., what humans see, has certain attributes, which are also visual sensations. These include color, visual texture, gloss, and transparency. In the ASTM definition the “spectral aspects” refer, of course, to color, and the remaining aspects (including the perceived shape, size, etc. of objects) are usually considered as “geometric aspects.” From these visual sensations observers get information about the environment and build their knowledge of the external world, which allows them to interact with the environment and with other individuals.

Historical Review of Studies on Visual Appearance

Under the heading of visual appearance a number of aspects are included such as color, texture, gloss, transparency, and also the shape and size of perceived objects or stimuli. The psychologist ► [David Katz](#) was probably the first to understand the various kinds of phenomena that always accompany the perception of color [3]. He describes different modes of color appearance: surface color, film color, volume color, specular colors, transparent colors, perception of luster, etc. By the middle of the twentieth century, Arthur Pope realized that, in order to define a color with accuracy, i.e., giving account for different modes in which a color may appear, more than the three usual variables (hue, lightness, and saturation) are needed [4, p. 28].

In 1953, the Committee on Colorimetry of the Optical Society of America published a book in which visual perception is classified into 11 “attributes of modes of appearance” [5]:

1. Brightness (or lightness)
2. Hue

3. Saturation
4. Size
5. Shape
6. Location
7. Flicker
8. Sparkle
9. Transparency
10. Glossiness
11. Luster

Lightness, hue, and saturation (attributes 1, 2, and 3) are the usual variables by which color sensations are described. Size, shape, and location (attributes 4, 5, and 6) are spatial categories of visual perception, which can be included in the category of spatial form. Flicker and sparkle (attributes 7 and 8) involve temporal variations in the perception of light. Finally, transparency, glossiness, and luster (attributes 9, 10, and 11) refer to the perception of different spatial distributions of light, aspects that will be included in the category of “cesia” to be described in the next section.

Richard Hunter proposed a classification of the geometric attributes of visual appearance [6]. He defined six different types of gloss and developed instruments for the measurement of some of these phenomena: ► [goniophotometers](#), diffuse-reflection and specular-reflection meters, ► [glossmeters](#), diffuse- and specular-transmission meters, etc. Before Hunter, August H. Pfund had already described an instrument for measuring gloss [7]. The American Society for Testing and Materials has standardized the measurement of various physical aspects related to visual appearance [8–11] and published a glossary of standard terminology [1].

Robert Hunt has developed a hypothesis for the measurement of color appearance under different conditions of illumination and observation, taking color as the central point of his study [12]. Sven Hesselgren observed that visual sensations, such as luster, reflection, and gloss, are not perceived as belonging to the color of an object but as something separate from color [13]. Ralph Evans, like Pope, also realized that the three variables normally used to define color are not enough to characterize color under different modes of

appearance [14]. Evans concluded that it would be necessary to define at least five variables. In these conceptions, the phenomena of appearance have always been considered as aspects that accompany color sensations.

Other researchers have studied, either partially or globally, the aspects of visual appearance related to the different modes of distribution of light in space. Fabio Metelli tackled the problem of *transparency* from a phenomenological point of view, establishing a clear difference between physical and perceived transparency and developing a model that predicts the conditions under which the perception of achromatic transparency is possible [15]. Osvaldo Da Pos, following Metelli, enlarged the model to study the perception of chromatic transparency [16].

John Hutchings studied the phenomenon of translucency and its importance for the visual qualities of food. In articles published in 1993 and 1995 he integrated all the aspects into a model of total appearance [17, 18]. However, this model is mainly oriented toward the analysis of food [19] and for this reason includes also nonvisual aspects of appearance such as smell, taste (flavor), and also texture as felt in the mouth, in addition to visual texture. Paul Green-Armytage proposed a three-dimensional model to organize what he called “qualities of surfaces” [20, 21]. Subsequently he introduced the term “*tincture*,” borrowed from heraldry, to encompass color, texture, and *cesia* [22]. Robert Sève tackled the specific problem of gloss [23]. Michael Brill used the model of Metelli to formulate a series of rules for the perception of chromatic translucency [24].

In 2006, a technical committee of the CIE (Commission Internationale de l’Eclairage, International Commission on Illumination) published a report to set a framework for the measurement of visual appearance [2].

A Classification of Visual Appearance into Four Categories

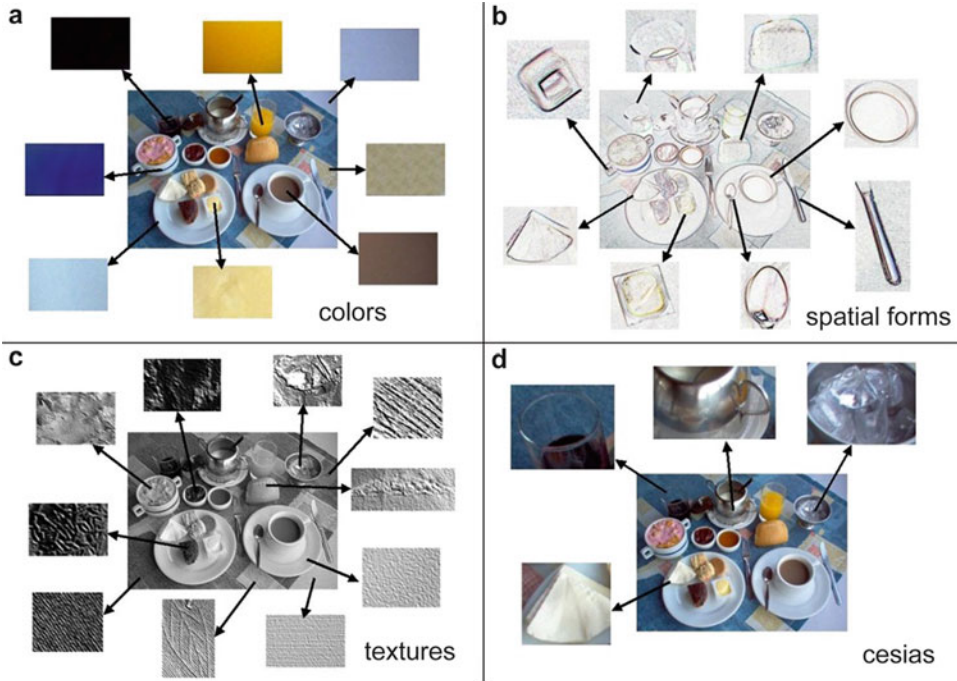
It is possible to analyze a given scene and try to understand how most humans separate and

categorize (by means of concepts and words) the main visual aspects involved. If the features shown in Fig. 1a are selected, people would agree on saying that they are different colors. If other features are selected, such as those shown in Fig. 1b, people will call them shapes (spatial forms). If the features selected in Fig. 1c are shown, people would call them textures. Finally, for the features selected in Fig. 1d the term *cesia* has been proposed.

Color can be defined as the aspect of visual appearance that results from the perception of different spectral compositions of light; it is what people are referring to when they say that some object or light stimulus is red, green, yellow, blue, black, white, gray, etc.

Spatial forms, or shapes, are built in consciousness from the perception of borders and their integration, so that they enclose or delimitate a portion of space. Spatial forms can be one dimensional (a straight line, a curved line, an irregular line, etc., which separate one portion of space from another), two dimensional (a triangle, square, circle, irregular surfaces, etc.), or three dimensional (a cube, a prism, a pyramid, a tetrahedron, a sphere, irregular volumes, etc.). They can have different shapes (as usually defined by the previous names), but they can also differ in size (length, surface, and volume) and proportion (defined by the relationships among the dimensions of the shape: length, width, and height).

Visual textures arise from the perception of sets of small elements that appear repeated in a surface or volume. The category of texture applies when these elements have no individual significance and are interpreted as making a whole, building a pattern that can be more or less regular or irregular. Visual textures can be distinguished and described by different variables or dimensions. These variables include density (the elements of the texture can be sparse, dense, or intermediate), predominant directionality (the elements can vary in direction or have no predominant directionality), and, of course, the shape and size of the relatively small elements whose repetition composes the texture. Visually perceived textures may be two dimensional (without relief) or three



Appearance, Fig. 1 The same scene with different visual features selected and categorized: (a) Different *colors*. (b) Different *spatial forms* or shapes that appear defined by their borders. (c) If attention is paid to certain regularities or repetition of elements on the objects, the notion of

texture can be applied. (d) Other visual characteristics such as the transparency of the glass, the white matte opacity of the paper towel, the metallic luster of the sugar pot, the glossy and semitransparent aspect of the ice cubes, and other similar ones have been given the term *cesia*

dimensional (with relief), the latter being also perceived by touch [25].

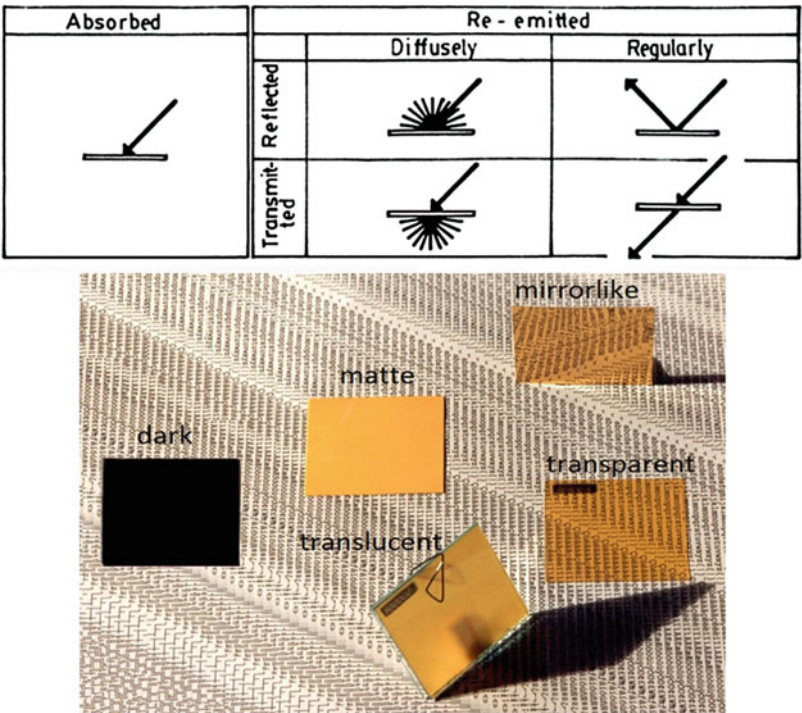
Beyond color, spatial forms, and textures, there are some other visual appearances that depend on the way in which differences in the spatial distribution of light are perceived. It is these spatial distributions of light that cause things to be seen as transparent or opaque, glossy or matte, or with intermediate appearances between these. Since there has not been a single global name or category to refer to these aspects, the neologism *cesia* has been coined for them.

Cesia

“Cesia” is a term invented around 1980 by César Jannello, to designate the perceptions resulting from the different spatial distributions of light. This concept was developed mainly by Caivano [26–29]. Cesia encompasses all the visual

sensations from transparency to translucency, matte opacity to mirror-like appearance (passing through glossy appearance), and lightness to darkness. Cesia are related to color in the sense that both are different aspects of the perception of light, and color deals also with lightness and darkness. But there are aspects that cannot be described by the three classical color dimensions of hue, lightness, and saturation. In this sense, cesia constitutes a category that can be added to the categories of color, visual texture, and spatial form in order to extend the possibilities of describing and classifying visual appearance. Three dimensions or variables can define cesia with respect to the perception of how light interacts with objects: permeability (from opaque to transparent), diffusivity (from regular to diffuse), and darkness (from light to dark). Light can be reflected or transmitted by objects, either diffusely or regularly, and it is the perception of these phenomena that is covered by the term “cesia.” For

Appearance,
Fig. 2 Cesia: processes followed by light falling on objects, and the way these stimuli are usually perceived



instance, a surface that reflects light in a diffuse way will normally be perceived to be matte. If a surface reflects light with a certain specular component it will be perceived as glossy or mirror like. If it transmits light diffusely it will be seen as translucent, and if it transmits light in a regular way it will be seen as transparent. Finally, a surface that absorbs most of the light will appear dark or black (Fig. 2).

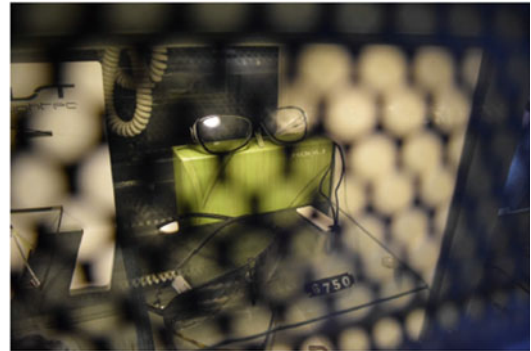
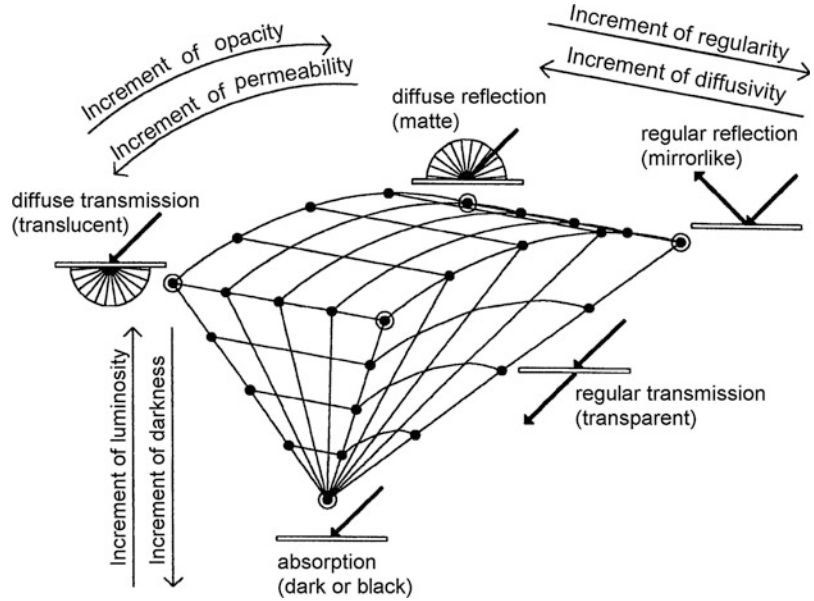
A three-dimensional model has been proposed to organize the three variables of cesia (permeability, diffusivity, and darkness). The model is a kind of inverted pyramid, with four vertices on the top for the sensations of matte appearance, mirror-like appearance, translucency, and transparency. The connections among these form scales with a variation of permeability (or the opposite, opacity) in one direction and scales with a variation of diffusivity (or the opposite, regularity) in a perpendicular direction. From any of these points, scales going down in lightness (or increasing darkness) connect to the point that is the place for black (Fig. 3).

In both color and cesia the relationship between stimulus and sensation is not fixed but

depends on three main factors – illumination, object, and observer – and is affected by other factors such as visual context, adaptation, contrast, etc. Some of these factors have been studied and exemplified with photographs by Maria P. Giglio [30]. To mention just an example: depending on the distance taken by the observer, an element can be perceived as more opaque or transparent, even when the physical arrangements of objects do not change (Fig. 4).

Color and cesia are closely connected because of their relationship with light; both are different aspects of the perception of light that contribute to the visual appearance of objects. The dimension of lightness/darkness is shared by color and cesia, being the link that connects both phenomena. Color and cesia interact thereby expanding the countless number of different visual appearances that humans are able to perceive. When Ludwig Wittgenstein, in his *Remarks on Color* [31], is concerned with the different types of white, yellow and golden, gray and silver, “black” mirrors, etc. and when he says, “Opaqueness is not a property of the white color; any more than

Appearance, Fig. 3 The three-dimensional model of cesias



Appearance, Fig. 4 A screened shutter that covers a shop window goes from near opacity to near transparency. When the observer is at a certain distance, the surface of the shutter looks more opaque than transparent and objects

behind it cannot be seen. When the observer gets closer, the small holes of the shutter tend to disappear, and the effect of transparency increases. (Photograph: M. Giglio)

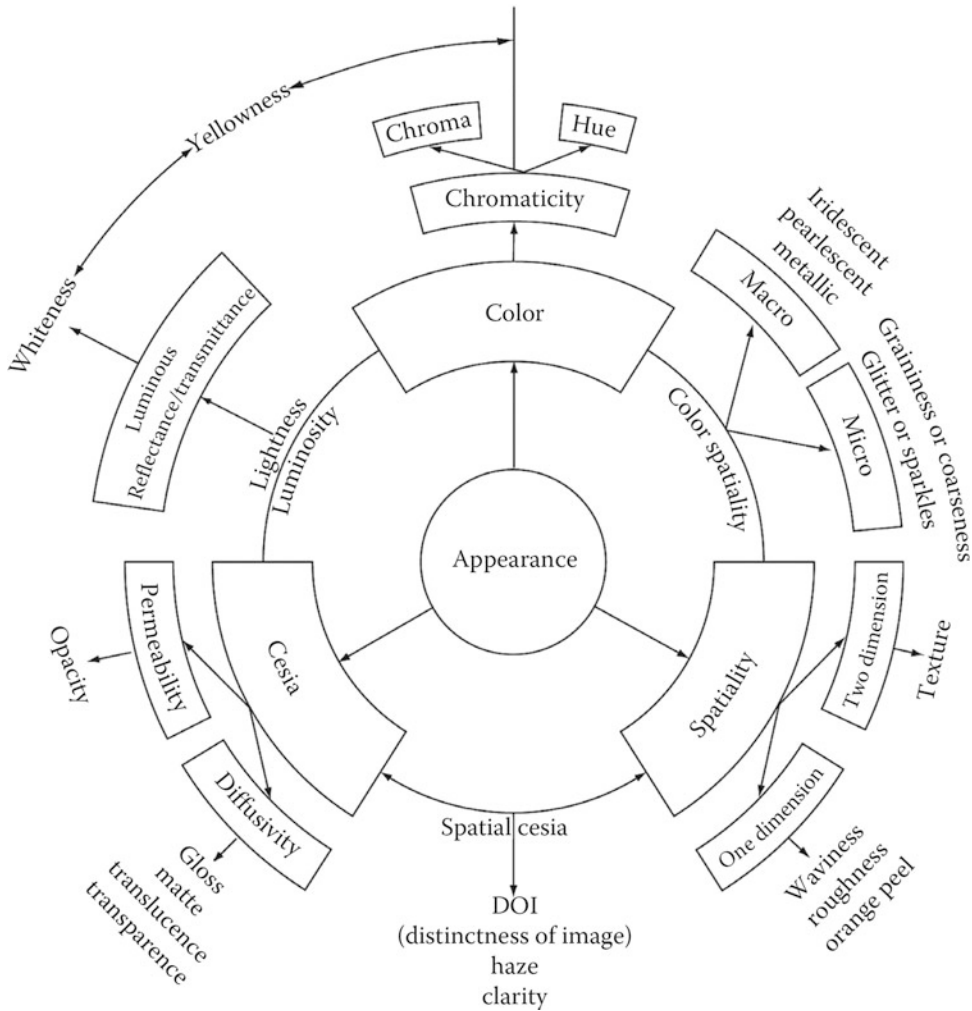
transparency is a property of the green,” he is dealing, in reality, with this aspect that we call cesia.

Models to Organize Different Aspects of Visual Appearance

Roberto Daniel Lozano, one of the members of the CIE technical committee 1-65, proposed a new approach to appearance characterization, with a model that classifies visual appearance

according to three main categories (color, cesia, and spatiality), which in turn includes subcategories and shows the interrelations and intermediate aspects among them (Fig. 5) [32].

More recently, Paul Green-Armytage proposed a model to link different modes and different aspects of appearance [33, 34]. The model shows relationships between the surface, volume, and illuminant modes of appearance as well as transparency, translucency, opacity, texture, gloss, metallic luster, and luminosity. Caivano identifies the five vertices of his inverted pyramid



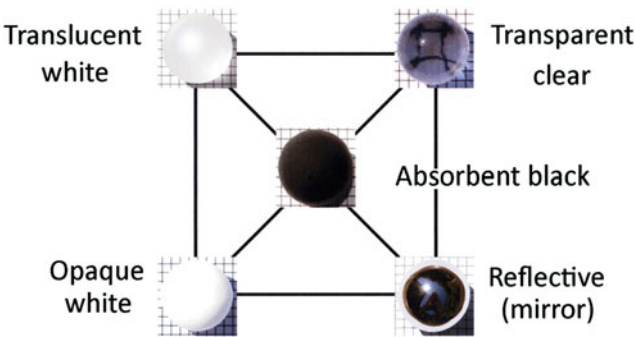
Appearance, Fig. 5 The classification of visual appearance proposed by Lozano. (Reproduced with permission of R. D. Lozano)

model of cesia (Fig. 3) as “primary sensations” [27]. These are opaque white, translucent white, transparent clear, mirror like, and black. Caivano’s three-dimensional inverted pyramid can be presented in two dimensions (Fig. 6).

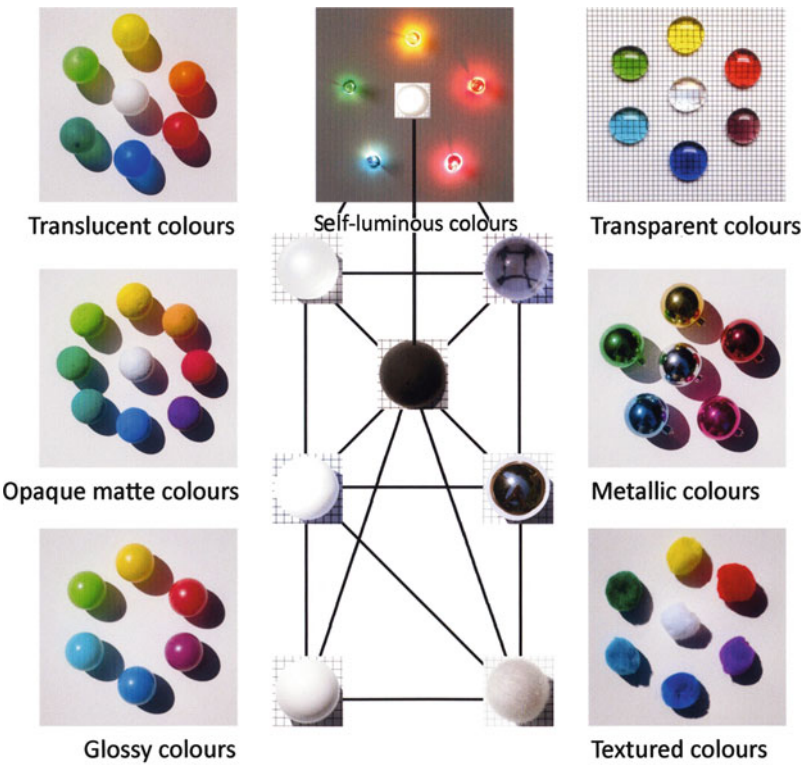
This two-dimensional representation of Caivano’s model is at the center of Green-Armytage’s model. Green-Armytage makes a distinction between gloss and metallic luster (mirror like), noting that scales can be constructed from a matte surface to surfaces that are increasingly glossy in one direction and from a matte surface to surfaces with increasing metallic luster in

another direction. He also notes that scales can be constructed with increasing texture from three different starting points: smooth matte, smooth gloss, and smooth metallic luster. Matte, glossy, textured, and metallic aspects of appearance are all characteristics of the surface mode of appearance. The volume mode can vary between clear (transparent) and cloudy (translucent). In the illuminant mode light sources can appear in varying degrees of brightness with a notional maximum that can be described as glare. To Caivano’s five primary sensations Green-Armytage adds three more: gloss white, textured white, and white

Appearance, Fig. 6 Two-dimensional representation of Caivano’s inverted pyramid model of cesia



Appearance, Fig. 7 Hue sequences in different aspects of appearance linked to Green-Armytage’s model



light glare. Each of the seven other primary sensations can be connected with scales to black, and these scales can, in turn, serve as the achromatic axes of order systems for colors that are opaque, translucent, transparent, and metallic (mirror like), as well as glossy, textured, and self-luminous. Green-Armytage’s model, with color circles to correspond with the different primary sensations, is shown in Fig. 7.

Increasing apparent intensities of illumination can be represented on lines which converge at

white light glare. The coherence of a light source, which affects appearances, can also be represented on the diagram. “Hard” light, as from direct sunlight and under which objects cast clearly defined shadows, is represented on the line connecting transparent clear to white light glare, and “soft” light, as from a heavily overcast sky and under which no shadows are cast, is represented on the line connecting translucent white to white light glare.

The various models described here can go some way to increasing awareness of the different aspects of appearance, and how they can be related, but no model can encompass the full experience of seeing the world in space and time.

Cross-References

- [Gloss Meter](#)
- [Goniophotometer](#)
- [Katz, David](#)
- [Tinture](#)
- [Transparency](#)

References

1. ASTM, American Society for Testing and Materials: Standard Terminology of Appearance. Standard E 284-01. ASTM, West Conshohocken (2001)
2. CIE, Commission Internationale de l'Eclairage, Technical Committee 1-65: CIE Technical Report: A Framework for the Measurement of Visual Appearance. CIE Central Bureau, Vienna (2006)
3. Katz, D.: Der Aufbau der Farbwelt, 2nd ed. of Die Erscheinungsweisen der Farben Und ihre Beeinflussung durch die individuelle Erfahrung. Johann Ambrosius Barth, Leipzig (1911/1930). English translation by MacLeod, R.B., Fox, C.W.: The World of Colour. Keagan Paul, Trench, Trubner, London (1935)
4. Pope, A.: The Language of Drawing and Painting. Harvard University Press, Cambridge, MA (1949)
5. Optical Society of America: Committee on Colorimetry: The Science of Color. Crowell, New York (1953) Chapter 2
6. Hunter, R.S.: The Measurement of Appearance. Wiley, New York (1975)
7. Pfund, A.H.: The measurement of gloss. J. Opt. Soc. Am. **20**, 23–26 (1930)
8. ASTM, American Society for Testing Materials: Standard Recommended Practice for Selection of Geometric Conditions for Measurement of Reflectance and Transmittance. Standard E, pp. 179–173. American Society for Testing and Materials, Philadelphia (1974)
9. ASTM, American Society for Testing and Materials: Standard Guide for Selection of Geometric Conditions for Measurement of Reflection and Transmission Properties of Materials. Standard E, p. 179. ASTM International (1990)
10. ASTM, American Society for Testing and Materials: Standard Test Method for Visual Evaluation of Gloss Differences Between Surfaces of Similar Appearance. Standard D, p. 4449. ASTM International (1990)
11. ASTM, American Society for Testing and Materials: ASTM Standards on Color and Appearance Measurement, 5th edn. ASTM, Philadelphia (1996)
12. Hunt, R.W.G.: Measurement of color appearance. J. Opt. Soc. Am. **55**(11), 1540–1551 (1965)
13. Hesselgren, S.: The Language of Architecture. Studentlitteratur, Lund (1967)
14. Evans, R.M.: The Perception of Color. Wiley, New York (1974)
15. Metelli, F.: The perception of transparency. Sci. Am. **230**(4), 90–98 (1974)
16. Da Pos, O.: Trasparenze. Icone Editrice, Padova (1990)
17. Hutchings, J.B.: The concept and philosophy of total appearance. In: Nemcsics, A., Schanda, J. (eds.) Colour 93, Proceedings of the 7th Congress of the International Colour Association, vol. C, pp. 55–59. Hungarian National Colour Committee, Budapest (1993)
18. Hutchings, J.B.: The continuity of colour, design, art, and science (I and II). Color. Res. Appl. **20**(5), 296–312 (1995)
19. Hutchings, J.B.: Food Color and Appearance, 2nd edn. Aspen Publishers, Gaithersburg (1999)
20. Green-Armytage, P.: Colour and other aspects of appearance. Spect. Newsl. Colour Soc, Aust. **6**(3), 1–11 (1992)
21. Green-Armytage, P.: Beyond colour. In: Nemcsics, A., Schanda, J. (eds.) Colour 93, Proceedings of the 7th Congress of the International Colour Association, vol. A, pp. 155–162. Hungarian National Colour Committee, Budapest (1993)
22. Green-Armytage, P.: Tinture – a new/old word for the appearance of things. J. Sch. Des. **2**, 16–23 (1993)
23. Sève, R.: Problems connected with the concept of gloss. Color. Res. Appl. **18**(4), 241–252 (1993)
24. Brill, M.H.: The perception of a colored translucent sheet on a background. Color. Res. Appl. **19**(1), 34–36 (1994)
25. Jannello, C.: Texture as a visual phenomenon. Archit. Des. **33**, 394–396 (1963)
26. Caivano, J.: Cesia: a system of visual signs complementing color. Color. Res. Appl. **16**, 258–268 (1991)
27. Caivano, J.: Appearance (cesia): construction of scales by means of spinning disks. Color. Res. Appl. **19**, 351–362 (1994)
28. Caivano, J.: Cesia: its relation to color in terms of the trichromatic theory. Die Farbe. **42**(1/3), 51–63 (1996)
29. Caivano, J.: An atlas of cesia with physical samples. In: AIC Color 97, Proceedings of the 8th Congress, vol. I, pp. 499–502. Color Science Association of Japan, Kyoto (1997)
30. Giglio, M.: Dynamics of color and cesia in the urban landscape: transience and perception of differences. In: Proceedings of the International Colour Association (AIC) Conference 2019, pp. 324–325. AIC, Newtown (2019)
31. Wittgenstein, L.: Bemerkungen über die Farben (1950–1951). English translation by McAlister, L.,

- Schättle, M.: *Remarks on Colour*. University of California Press, Berkeley (1977)
32. Lozano, R.D.: A new approach to appearance characterization. *Color. Res. Appl.* **31**, 164–167 (2006)
 33. Green-Armytage, P.: A model to link different modes and different aspects of appearance. In: *AIC 2011 Interaction of colour and light Proceedings of the Midterm Meeting of the International Color Association*, pp. 165–168. Pro/colore, Zurich (2011)
 34. Green-Armytage, P.: More than colour – dimensions of light and appearance. *J. Int. Color. Assoc.* **17**, 1–27 (2017)

Arc Lamps

► Carbon Arc Lamp

Architectural Lighting

Kevin Mansfield and Peter Raynham
University College of London Institute for
Environmental Design and Engineering,
London, UK

Synonyms

[Lighting design](#)

Definition

Architectural lighting design is the design and implementation of daylight and electric lighting provision in urban and architectural environments. This process requires conceptual thinking, design development, and illuminating engineering to produce lit environments for people's work and leisure.

Lighting Quality

The built environment in which people live and work is composed of a variety of building types. In a shopping mall, retail units are expected to be

colorful and attractive. In an art gallery, it is necessary to be able to appreciate the color and detail in a painting, and when walking in the city, people expect to remain safe and secure. All these impressions can be dramatically affected by the lighting. Boyce has suggested that *bad* lighting “does not allow you to see what you need to see, quickly and easily and/or causes you discomfort” [1]. If the lighting designer overcomes these problems, then they can produce – at best – *indifferent* lighting. Only when the configuration of the windows and surfaces or the characteristics of the lighting installation “lift the spirits”, can the environment be described as *good* lighting.

The task of the lighting designer or a lighting engineer is to determine the character of a lit space, whether it be a street, a townscape, or a room, and then to arrange the sources of light, be they daylight or electric light, to achieve this desired character.

Objectives of Design

The purpose of lighting is to provide for the needs of people who are going to use the built environment that is being created, and any person designing the lighting will wish to create a quality lit environment. However, this raises the question about what constitutes good quality lighting? To answer that question, the needs of the people who are going to use the building that is to be lit require consideration. In general terms, the users of a building are going to want lighting so that they can see things easily, they will require the lighting to set the correct ambience, and if they are going to remain in the building for long periods, they may well benefit from the biological impact that the lighting may have on their endocrine systems. These topics are covered in more detail in the following sections.

Light for Seeing

Before considering what sort of lighting is required, it is first essential to consider what the occupants of a building need to see and what the consequences are of them not being able to see. Consider two cases: a surgeon undertaking heart

surgery and an office junior doing a filing task. Both of them need to be able to see small objects – in the case of the surgeon, this may be small blood vessels; whereas, for the office junior, he has to read the small text on the documents he is working with. However, the office junior is normally dealing with high contrast printed text, while the surgeon has the problem of distinguishing objects that are very similar in color and reflectance. Moreover, if the surgeon makes a mistake, it could kill a patient, but an error in filing can be easily rectified. Thus, it is normal to have more than ten times as much light available in an operating theatre compared to an office.

Selecting an appropriate level of lighting for a given visual task is bound to be complex as a number of factors need to be considered. These include the complexity of the visual task, the visual acuity of the person performing the task, the length of time for which the task is performed, and the consequences of any mistakes. These factors need to be set against the cost of providing the lighting and the other negative consequences of providing lighting such as energy use and associated greenhouse gas emissions.

To help lighting designers in the selection of appropriate lighting conditions, there are a number of codes and standards available which provide recommendations for the lighting conditions considered necessary for a whole range of different tasks. These documents have evolved over a number of years and in general provide a consensus view of the lighting needs in a typical installation. However, the lighting requirements for every installation are potentially different and so the guidance provided by codes and standards should usually be regarded as a starting point from which the design may be developed; however, there are occasionally situations where the national regulations insist that particular illuminance values are used in given circumstances. Good examples of such codes are the *IES Handbook* [2], the *SLL Code for Lighting* [3], and *ISO 8995-1:2002* [4].

While considering the required illuminance necessary for a particular task, it is also important to consider other factors such as freedom from glare and the necessary color properties of the

light sources used. Most codes and standards also give advice on these issues.

Setting the Ambience

Every lighting installation sends a message to the users of the building; it is the art of good lighting design to ensure that this message complements the message sent by the architecture and that the overall message is appropriate to the use of the building. Consider the lighting in a supermarket and compare it with the lighting in an up-market boutique. In many ways, the lighting has a similar job to do – it must show off the goods in the store so that people choose to buy them. However, the style of the lighting is radically different as supermarket style lighting in the boutique would send entirely the wrong message to potential customers.

There is no simple way of providing lighting that perfectly provides the correct message for a given environment. There have been many studies into how to characterize the lit environment of places that are considered to be well lit, and in general, the results have been patchy and limited in validity to the environment in which they were assessed. Section “[A Framework for Design](#)” gives some basic advice, but it is always going to involve a series of complex subjective judgments to get the lighting right and the reliance on simple codes will not suffice.

Biological Impact

The nonvisual or biological effects of light have their biggest impact on the regulation of the human sleep/wake cycle. Exposure to high levels of light may be used to shift the time of peoples’ body clocks and may also be used to boost the alertness of people. These biological effects happen naturally to people in an outdoor environment; however, in many interiors, there is not sufficient light to provide the necessary stimulus to the human endocrine system.

The most logical way to resolve the problem is to provide interiors with good levels of daylight. Light sources may be used, biased toward the blue and blue/green regions of the visible spectrum (within acceptable safety limits), if the objective of the design is to promote the nonimage-forming impacts of light [5].

A Framework for Design

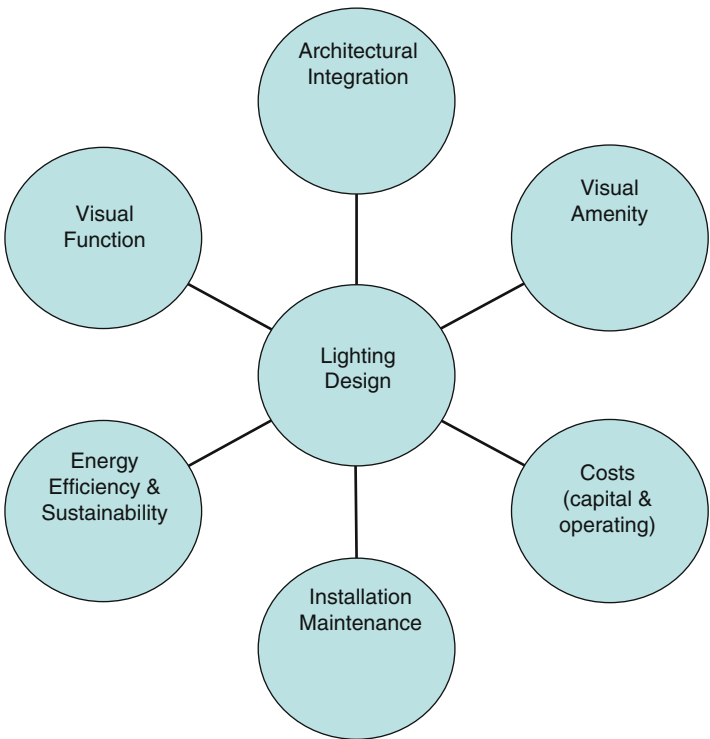
The objective for anyone trying to design lighting is to provide good quality lighting. The construction of a new building or the refurbishment of an existing one comprises a series of construction stages from design through construction to the final use of the building or area. Lighting should be designed by someone who has the ability to make the necessary aesthetic judgments about how the lighting will impact on the space being lit while at the same time engineering the lighting equipment into a functioning system. Therefore, the lighting designer or lighting engineer can make a particular contribution or intervention within the *design* phase. Here, a lighting design concept is proposed where outline proposals for the arrangement of structural and building services are made. The design is then developed in detail to decide on the relationship between the electric lighting systems and the daylighting, and finally a technical specification is drawn up to coordinate the various components and architectural elements.

At each of these stages, the designer will be considering a variety of factors. Veitch has suggested that designers should consider the well-being of the individual user, the integration with the architecture, and the economics of the lighting installation [6]. Loe has proposed a useful variation in a framework for design [7], which comprises the following elements (Fig. 1).

Lighting for visual function: The lighting should enable users to undertake visual tasks without visual discomfort. Recommendations are made concerning the level of light on the task (illuminance) and the uniformity of light on the task. It is also important to prevent unwanted brightness in the visual field (the control of glare).

Lighting for visual amenity: People prefer an architectural room or urban space to contain a degree of *visual lightness* (a sense of lightness and airiness) and *visual interest* (a variation in light and shade). The color appearance of the space should be appropriate for the application, e.g., warm and inviting or cool and professional. The objects and people in the space need to be lit pleasingly.

Architectural Lighting,
Fig. 1 A framework for lighting design (after Loe). The relative weight given to each element of the framework depends upon the requirements of the lighting design



Energy efficiency and sustainability in lighting: The lighting installation must contribute to sustainability by maximizing the use of daylighting and ensuring that energy efficient light sources and light fittings are specified. Reducing energy consumption is a matter of reducing the power load of the lighting components and the hours of use. This can be achieved by the effective use of lighting control systems (switching, dimming, and linking to daylight levels).

Lighting installation maintenance: The lighting designer or engineer must specify a lighting installation that can be easily maintained by the client or user. It should be easy to clean and relamp the fittings to ensure that the installation continues to deliver the appropriate lighting. Poor maintenance of an installation will lead to some of the lighting points failing and this sends the message that no one cares about the space.

Lighting costs: The cost of the installation needs to be controlled and a distinction is made between *capital* costs (for the original lighting equipment) and *running* costs (the cost of operating the lighting system as designed).

Architectural integration: An understanding of the intent of the architect and the character of the space in terms of dimensions, finishes, textures, and materials is essential. Only then can electric lighting systems be properly integrated into the fabric of the building to produce the appropriate lighting effect.

These factors are described in greater detail in the *SLL Lighting Handbook* [8]. It should be noted that they are occasionally legal requirements for lighting in certain areas.

Daylighting

It is useful to think of the elements of the city and the town and the constituent buildings as being under the vast dome of the sky. In many climates, there will be days in which this dome is covered by a thick layer of cloud – the fully overcast sky, which might be termed the *diffuse* component. As the conditions change, the clouds begin to thin and the sun emerges. This source has a very different character – a sharp beam and the ability to

create shadows – this may be called the *direct* component. These two components of daylighting are varying continually throughout the day and throughout the year.

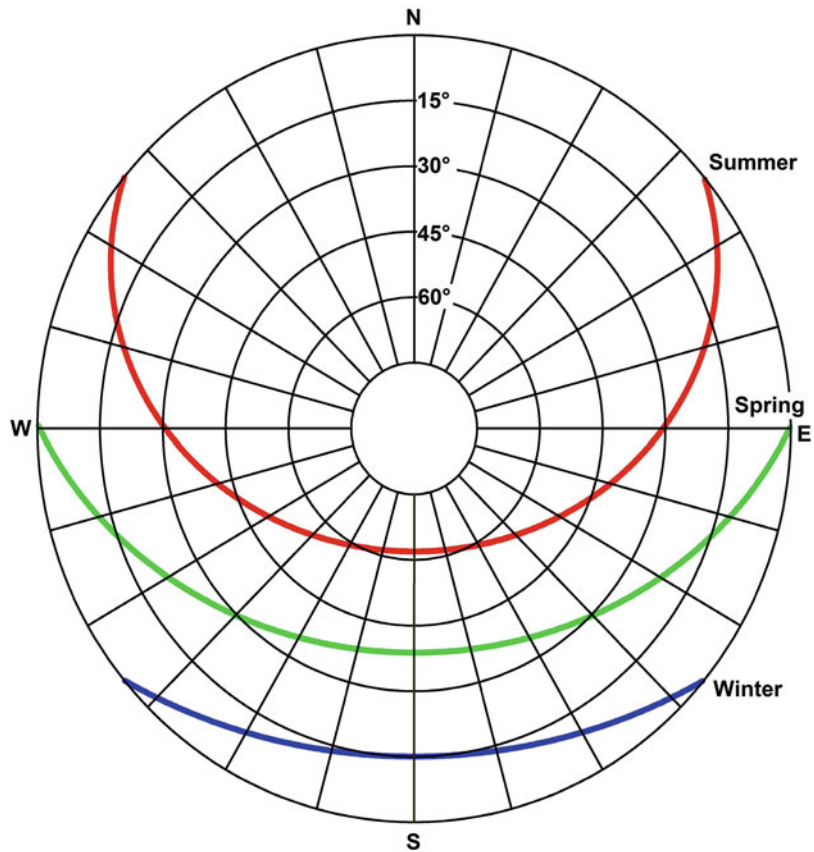
The lighting designer works at different scales of analysis. For example, the designer might seek to optimize the spacing between apartment blocks to ensure that certain apartments are not restricted in the amount of light that they receive. The designer might also seek to ensure that courtyards or plazas receive enough sunlight or are shaded in certain climates. Simulation software has the capability to demonstrate the effect of sun position on buildings. Stereographic sun paths are available to show the passage of the sun in the sky (Fig. 2).

The lighting designer can also work in conjunction with the architect to ensure that the individual rooms and spaces within a building receive sufficient daylighting. The traditional way to pierce the solid wall of a building is by a vertical window. The clerestory window and the rooflight provide ways of allowing daylight to enter the building to provide a different character to the space. This is because they are facing different parts of the sky. The atrium is a useful device, historically and in contemporary buildings, to let natural light deep into the building. It is also possible to “borrow” light from one space to another using transparent partitions.

Contemporary buildings have a framed steel or concrete structure in which the structure defines the architecture. The introduction of daylight into the building thus becomes a result of the structural decisions that have been made. This is the province of facade engineering whereby the transparent partitions that clothe the structure are then subject to varying architectural treatment in terms of the choice and specification of the glazing (solar control or electrochromic materials) and whether shading is to be added (louvres and *brise-soleil*). It is at this point that daylight availability is estimated by calculation of the daylight factor or methods of calculation of cumulative daylight exposure. The European standard for daylight in buildings [9] provides guidance on the necessary qualities of daylight to necessary to provide a successfully daylit space.

Architectural Lighting,

Fig. 2 A plot of the sun's position in the sky for London (51°N Latitude) denoted by its azimuth angle in plan and altitude above the horizon



When designing the daylight in a building, it is also important to consider the quality of the view out (and in) that the windows may give.

Electric Lighting

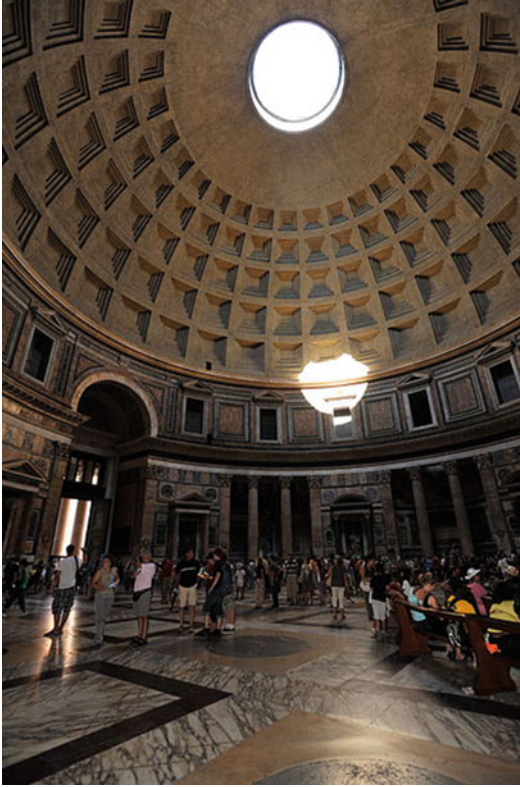
There are a plethora of new technologies for lighting, e.g., light-emitting diodes (LEDs), and a variety of lamp technologies including high-pressure discharge lamps, fluorescent lamps, and incandescent lamps. The designer has the ability to control both the intensity and the color of the light that is emitted from these sources.

Consider the public square in the town or city. An appropriate choice of lamp on an amenity lighting column can attractively reveal the features on the faces of fellow street users to ensure safety and security, and the appropriate color choice can dynamically reveal the city.

The architectural façade can take a variety of forms: neoclassical, turn of the century, international modern, or contemporary. It is the task of the lighting designer to reveal the characteristics of the architecture without distortion, to complement and enhance the architectural detail. It is important to ensure also that each architectural element is read as a whole – from the base to the roof. Again the judicious choice of color and intensity of the light source in conjunction with an appropriate lighting fitting can ensure that the architectural intention is maintained.

Moving indoors the choice of light source remains critical in terms of its color, its life, and its efficiency. Such sources are housed within light fittings (*luminaires*) – used individually or more typically in arrays to provide layers of lighting or inbuilt into the architecture.

An individual luminaire can be used to accent individual features, and arrays of luminaires



Architectural Lighting, Fig. 3 The Pantheon in Rome

(linear and modular in form) can be arranged within the various architectural spaces to provide lighting for tasks and to reveal the architecture of the space with its characteristic finishes and materials. Care is taken by the designer to ensure that the user of the space is not subject to unwanted glare from the fittings and that the overall brightness of a sequence of spaces from one to the next is kept within reasonable bounds (*adaptation*).

The design of an individual luminaire is a complex process in itself. It must be designed to house a particular light source which might be spherical, linear, or elliptical. The flux from the source must be possibly redirected into the required direction by the use of a reflector and then the beam from the lamp must be further refined by optical elements such as diffusers, lenses, or louvres. The fitting must be electrically safe, must not overheat, and be mechanically strong – these are common requirements even though the physical form of the luminaire might be quite different cf. a downlighter in a conference center to a street lighting bollard. The designer must also consider the appearance of the fitting when it is lit and when it is unlit and ensure that

**Architectural Lighting,
Fig. 4** The City of Lyon at
night



Architectural Lighting,
Fig. 5 A luminous media
 façade



A

the fittings integrate well with the architectural character of the space.

Endpiece

Architectural lighting has great power. Think of the dramatic daylighting of the Pantheon in Rome (Fig. 3) or the harmonious integration of the lighting of the city of Lyon at night (Fig. 4) or the dynamic lighting of a modern luminous media façade (Fig. 5). But with that power comes responsibility – responsibility that can be accepted by the designer in designing lighting that promotes human comfort and satisfaction.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Lighting Controls](#)

References

1. Boyce, P.R.: Lighting quality: the unanswered questions. Paper presented at the first CIE symposium on lighting quality, Ontario, Canada, 9–10 May 1998
2. Illuminating Engineering Society: Lighting Handbook. Illuminating Engineering Society, New York (2011)

3. Society of Light and Lighting: SLL Code for Lighting, London (2012)
4. ISO 8995-1:2002 (CIE S008/E:2001) Lighting of Work Places Part 1: Indoor
5. Lucas, R.J.: Measuring and using light in the melanopsin age. *Trends Neurosci.* **37**(1), 1–9 (2014). <https://doi.org/10.1016/j.tins.2013.10.004>
6. Veitch, J.A.: Commentary: on unanswered questions. Paper presented at the first CIE symposium on lighting quality, Ontario, Canada, 9–10 May 1998
7. Loe, D.L.: Lighting quality – an exploration. Paper presented at the first CIE symposium on lighting quality, Ontario, Canada, 9–10 May 1998
8. Society of Light and Lighting: SLL Lighting Handbook, London (2018)
9. EN 17037: 2019 Daylighting in Buildings

Art and Fashion Color Design

Larissa Noury
 Colour-Space-Culture, Paris, France

Synonyms

[Color harmonies for fashion design](#); [Modern art and fashion design](#); [Painted dress](#); [Tactile painting and haute couture](#)

Definition

Art and fashion color design concerns the harmonies of color associations of textile and other materials used for a couturier's aesthetic project. It is an illustration of collaboration of artists and couturiers on the harmonies of color associations in order to create a personal style. They were more closely tied at the turn of the twentieth century than they are today. Artists did not see the difference between creating an original work of art, such as a painting, and designing a textile pattern that would be reproduced many times over. Each was a valid creative act in their eyes. There are a lot of vivid illustrations of the centuries-long love affair between fashion and art of color. Couturiers are past masters at capturing the contemporary *zeitgeist* in their designs, while artists have frequently used clothing as a way to give all-round expression to their aesthetic ideas.

The Role of Art and Fashion

Beginning from the nineteenth century, major changes emerged both in the role of fashion and in the place of art in society. Growing affluence and new social structures gradually turned art, color, and fashion into ways of expressing personal taste and identity. This section presents a historic panorama of the color in the fashion design during the twentieth century, highlighting certain symbolic movements such as the Art Nouveau, the Russian avant-garde, the modernism, pop art, or the kinetic art. It lists the color's harmonies for modern fashion design and describes the tools, ranges, palettes, and techniques which allow personalizing a dress with fantasy and subtlety.

Art has often been a major source of inspiration for dress designers of the twentieth century. Some creations by Liberty, Sonia Delaunay, Liza Schiaparelli, Coco Chanel, Yves Saint Laurent, Givenchy, and Jean-Paul Gauthier can be mentioned. In their practice, the art of color is conveyed to fashion. The designers paraphrase the masters of painting. This connection between art

and fashion is particularly obvious starting in the late nineteenth century.

Fashion and Identity of the Color Palette

Belle Epoque

The artistic principles of the Art Nouveau combine the personal style of expression where the sinuosity of the shape is harmonized well with the mosaic of contrasted colors. The synthesis of practical considerations and of aesthetic natural forms and lines is characterized by color palettes of a great richness. The modern time influenced a lot the development of the fashion design and produced profusion of the ► [polychromy](#) in it. The large selection of paintings by Giovanni Boldini, Auguste Toulmouche, James Tissot, Jean Beraud, and Alfred Stevens – visual artists who reflected the best on the fashion style of the Belle Epoque – is an example of what could be called fashion inspiration in fine arts. Artworks were created in a realistic manner and have not even a touch of stylization typical of modernist art. Figures depicted in these paintings are precise and realistic. However, it would be fair to say that style itself is the subject matter instead of a particular model. Fashion is the main theme and inspiration for these paintings.

Beautiful symbolic women's dresses by famous Austrian artist Gustav Klimt followed, in which he introduced his frescoes and paintings of portrait. Klimt himself drew the blue caftan, a long tunic of oriental style, and many other dresses for Flöge Emilie, his companion for many years, who owned a fashion house in Vienna.

After a visit of a workshop in Vienna, the avant-garde fashion designer Paul Poiret brought the idea of mixing art and fashion to Paris. He opened a Martine school in 1911, a place that was also attractive for artists. He then employed Parisian artists such as Lepape, Ibibe, and Erte on fashion illustrations. He employed the artist Raoul Duffy to design fabric prints and to invent tissues. He went to art galleries and showed his artistic sensibilities by preferring Impressionist paintings at a time when they were new and

unappreciated by the public. Poiret became very interested in modern art and said, “I have always liked painters. It seems to me that we are in the same trade and that they are my colleagues.” The couturier considered himself as an artist first.

Russian Constructivism and Suprematism

The artists of Russian avant-garde in 1915–1935 initiated probably the most intensive and creative art and architectural movement of the twentieth century and became a significant source of any art movement since then. The constructivists defined the chromatic surfaces as fundamental colored elements where the straight lines, the rectangular forms, and the principal colors (yellow, red, blue, black, and white) are used to make a unified composition. The color interdependence simultaneously with the community of ideas formed by the uniformity of geometric contours and the absence of force of gravity, the balance of the unit, and the dynamism of the parts, all form a kind of representation of an ideal cosmos.

Varvara Stepanova, Alexandre Rodchenko, Liubov Popova, and others show their creativity thorough re-energizing new forms and meaning in art and dress design. The key fragments of Russian revolutionary creativity still glow like radium, living on futuristic art and design into the imaginations of some most influential couturiers of the twentieth and twenty-first centuries. Christian Dior *haute couture* in 2002 was inspired by the color ► [palette](#) of Malevich painting.

Simultaneous Contrast and Sonia Delaunay

In the 1920s, abstract painting inspired a variety of fabric designs, mainly based on the simultaneous contrast effect, by the successful designer and artist Sonia Delaunay [1]. Married to Robert Delaunay and friend of artists such as Mondrian, Arp, Vantongerloo, and Kandinsky, she was a member of the contemporary artistic avant-garde in her own right. It was her own abstract paintings that she translated into rhythmic designs composed of squares, lines, circles, diagonals, and color planes. In all, Delaunay created over 2,000 of these fabric designs, around 200 of them produced especially for the fashion house Metz & Co in Amsterdam.

Surrealism and Elsa Schiaparelli

Fashion designer Elsa Schiaparelli, Coco Chanel’s main rival in the 1920s and 1930s, produced clothing and hats heavily influenced by Surrealism [2]. Her sweaters incorporating knitted ties or sailor collars were a sensation, and she worked in close cooperation with artists such as Salvador Dalí and Jean Cocteau. An example of her work with Dalí is her famous lobster dress.

Coco Chanel and Japanese Art

The designer’s passionate interests inspired Coco Chanel’s fashions [3]. Her apartment and her clothing followed her favorite color ► [palette](#), shades of beige, black, and white. Elements from her art collection and theatrical interests likewise provided themes for her collections. The ornament of the dress, in both pattern and color palette, resembles the Asian lacquered screens which the designer loved and collected. In convergence with the Art Deco line, the modernist impulse was married with pure form and Japanese potential.

Neoplasticism and the Bauhaus

Piet Mondrian changed the face of modern art. His influence extends to painting, sculpture, graphic design, and fashion. In search of plastic harmony, he introduced a universal language of shapes and ► [primary colors](#) that goes beyond the painting; Mondrian was the central and most famous figure of the De Stijl movement. This style was baptized as neoplasticism and intended to achieve real objectivity by releasing the work of art from its dependence on the momentary individual perception and temperament of the artist. Yves Saint Laurent has created his famous dress with Mondrian’s color composition. The color in fashion design was developed by the school of Bauhaus and was also very enriching and interesting [4]. The relationship between form and color within the framework of visual perception was defined starting from the color theories of Kandinsky, Klee, Itten, and Albers in this school.

Op Art, Minimalism, and Pop Art in Fashion

Since the Second World War, there has also been frequent interaction between art and fashion. In the 1950s, Karel Appel produced signed fabric

designs; in the 1960s, the Bijenkorf department store sold dresses inspired by pop art, and since then the emergence of minimalist art has given rise to a widespread taste for sober, often asymmetrical designs. For instance, Jean-Charles de Castelbajac was inspired by Andy Warhol and his pop art *Campbell's Soup* painting. Victor Vasarely, Yvaral, and other representatives of optical art, as well as Friedensreich Hundertwasser's art, continue to influence modern fashion design and modern life [5].

In 2007, Christian Dior designed a unique piece, hand painted and enhanced with spectacular embroidery, Manteau Suzurka-San. Dior *haute couture* was inspired by *The Great Wave of Kanagawa*, emblematic work of the Japanese artist Hokusai. Focusing on the relationship between art and creations of the house Dior, it is possible to say that the original works have been in one way or another influenced by different artists. Certain highlighted artworks influenced not only the collections *haute couture* and *pret-à-porter* but also the unique world of fine jewelry, perfumes, and accessories. Because they reflect light and velocities, colors and shapes of the paintings of Sonia Delaunay emphasized the dancing model dressed in her parade by Ungaro Fall-Winter 2003–2004. Yves Saint Laurent revolutionized fashion and gave women the freedom of movement that has inspired artists, poets, and painters. “The profession needs an artist to exist,” he said. He loved painting and painters such as Matisse, Mondrian, Braque, Picasso, and Van Gogh. Saint Laurent began in 1988 with Georges Braque, whose famous birds seem to fly stuck to the bride's dress. Then, he designed a jacket inspired by *Iris* of Van Gogh. It took 800 hours to sew the whole Van Gogh. Flakes, tubes, seed beads, and ribbons were all embroidered by hand to make the effects and lighting of the canvas. Yves Saint Laurent influenced in some way other couturiers. For instance, stylist Erin Fee has been inspired by flowered textile design and some kind of “camouflage” for his fashion collection. Jean-Charles de Castelbajac, on the other hand, brings a joyful color palette from BD art graphics. Nowadays,

fashion designs have increasingly been regarded as autonomous works of art. Some creations by designers such as Viktor & Rolf now go so far in that direction as to be scarcely wearable at all. Viktor & Rolf spread their cutup couture in the magazine *Dazed and Confused*.

Tactile Color and Haute Couture

Nowadays, the collaboration between artists and fashion designers continues. The personal artist's inspiration comes from tactile painting. Jean Marie Pujol, couturier who worked with Dior and Yves Saint Laurent, designed several dresses to be painted by Larissa Noury as a means to perpetuate the art-and-fashion marriage [6]. With this personal style, a series of hand-painted dresses were created and presented during the exhibition at Faubourg's gallery in Paris, amid boutiques of Christian Dior, Jean Paul Gauthier, Chanel, and other creators of fashion.

Cross-References

- [Color Contrast](#)
- [Color Harmony](#)
- [Palette](#)
- [Polychromy](#)
- [Primary Colors](#)

References

1. de Leeuw-de Monti, M., Timmer, P.: *Colour Moves: Art and Fashion by Sonia Delaunay*. Thames & Hudson, London (2011)
2. Martin, R.: *Fashion & Surrealism*, reprint edn. Rizzoli International Publications, New York (1996)
3. Degunst, S.: *Coco Chanel: Citations*. Éditions du Huitième Jour, Paris (2008)
4. Fiedler, J., Feierabend, P.: *Bauhaus*. Editions Konemann Verlag, Cologne (2000)
5. Popper, F.: *From Technological to Virtual Art*. Leonardo Books. MIT Press, Cambridge (2007)
6. Pujol, J.-M.: *Haute Couture*. Éditions du Palais, Paris (2011)

Artistic Rendering

► Non-photorealistic Rendering

Assimilation

Simone Gori
Department of Human and Social Sciences,
University of Bergamo, Bergamo, Italy

Synonyms

Bezold–Brücke effect; von Bezold spreading effect

Definition

Color assimilation is by some considered to be the opposite of the color contrast: the color is perceived in the direction of the hue of the surrounding color, whereas in color contrast the perceived color tends to be the complementary color. Assimilation can be experienced also in grayscale stimuli, and it is called lightness assimilation or brightness assimilation. In this case a gray is perceived lighter if it is close to a light object and darker if it is close to a darker object which is, again, the opposite of the simultaneous contrast.

Introduction

According to a naive physicalism, the physical stimulus should present a direct relationship with its mental correlate. Visual illusions (e.g., [1–3]) indicate, however, that physical manipulations of a stimulus do not directly determine the perceptual experience [4]. The existence of visual illusions has been already reported by ancient philosophers, and they were commonly considered counterintuitive singularities [5], which

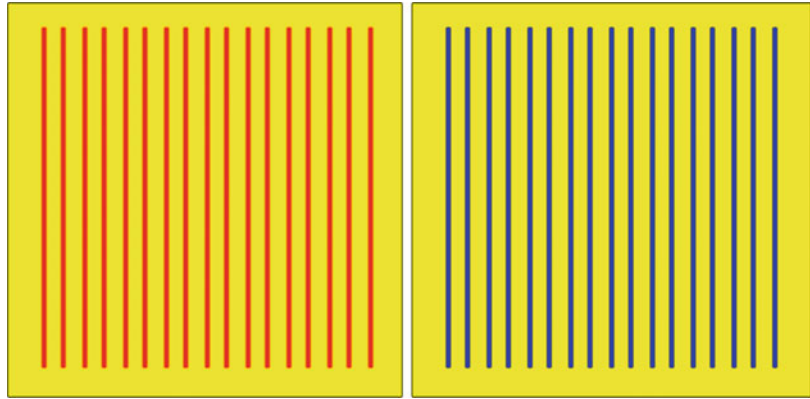
demonstrate the active work of the sensorial system involved in the stimulus processing [6, 7]. It was proved that also non-human animals can experience visual illusions [8, 9]. An important sub-group of this illusion is the brightness illusion, and this also shows a similar counterpart in colored stimuli. It is often assumed that perceiving a surface as a source of light depends just on its physical radiant emission. However, the Persian natural philosopher Ibn Al-Haytham (circa 965–1040 AD), known as Alhazen, stressed the subjective nature of color sensation and argued in his description of the simultaneous contrast that color appearance is partly due to a mental process [10].

Simultaneous contrast can be described as follows: a gray target surrounded by a bright inducer that appears darker than its physical value. If the same gray target will be surrounded by a dark inducer, the results will be the opposite: the target will be perceived lighter than its physical value. The simultaneous contrast can be enhanced by blurring the boundaries of the inducer [11–14]. The simultaneous contrast is observed also in color stimuli where the color of the target tends to the hue of the opposite color of the inducer.

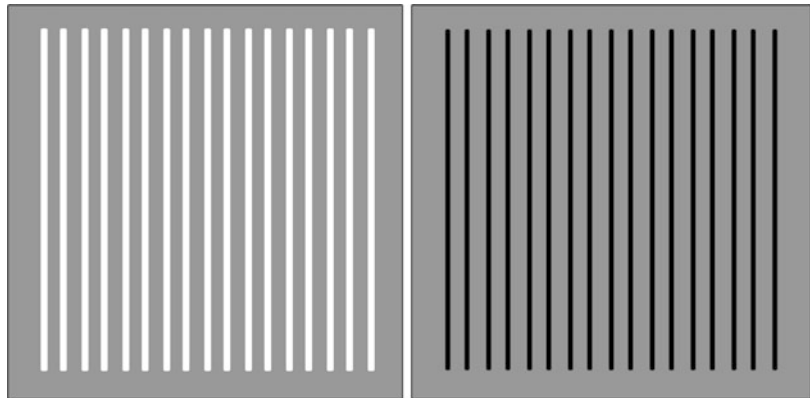
Assimilation Description

The brightness assimilation can be considered the opposite of the simultaneous contrast because under specific conditions a gray target will appear lighter when bordered by a brighter inducer and vice versa. The color assimilation is a similar phenomenon in which a colored target tends to be perceived as trending in hue toward that of the inducer color. Figure 1 shows a typical color assimilation display: the yellow background on the left is perceived slightly reddish, while the yellow background on the right is perceived as slightly bluish; however, the two yellows have identical reflectance functions. Figure 2 shows brightness assimilation: the gray background on the left is perceived lighter, while the gray background on the right is commonly perceived

Assimilation, Fig. 1 The color assimilation display: the yellow background on the left is perceived reddish, while the yellow background on the right is perceived bluish. The two yellows are physically the exact the same hue



Assimilation, Fig. 2 The brightness assimilation display: the gray background on the left is perceived lighter, while the gray background on the right is commonly perceived darker. The two gray values present the exact same luminance



darker; however, the two gray values have the exact same luminance. von Bezold [15] probably was the first to describe this effect in 1874. Based on the seminal observations by von Bezold, it seems that reducing the inducers' size and increasing their number (which implies increasing their spatial frequency and density together) will lead to a shift from simultaneous contrast to assimilation. Consequently, it seems clear that the physical variables responsible for this shift are the spatial frequency and the reflectance of the inducers. It is known, for example, that those two variables also affected the size perception [16]. Recently, it was demonstrated that even the perception of beauty is influenced by the spatial frequency [17]. However, the questions that remain open are: At which level of the brain processing are those variables interpreted in this specific way? And how does the visual system transform those variables in the final percept?

Possible Explanations for Assimilation

One approach to explain the assimilation is that bottom-up, peripheral mechanisms are sufficient to produce this percept. Assimilation could be indeed produced by the fact that the retinal input can be imagined as a blurry image; consequently, if the visual system will consider the retinal input of a display like the one in Figs. 1 or 2, the final percept will be in the direction of assimilation. However, considering how important is the cortical process in the final percept, it is hard to believe that at least in typically developed individuals the cortex will simply "accept" the retinal signal without any post processing in order to provide the final percept. Other proposals that consider the assimilation exclusively a bottom-up process are based on local averaging of luminance within large neurons' receptive fields. The receptive fields are small in the fovea and larger in the

periphery, and they also increase their size going up in the cortical hierarchy [18]. Specifically, two possible mechanisms have been proposed: a neuronal spatial integration [19] or neural weighted averages [20], suggesting that the primary anatomical site for assimilation could be spatially close but still outside V1. Both mechanisms may result in assimilation when the physical stimulus is a similar pattern to the one shown in Figs. 1 and 2. However, some predictions based on this explanation fell short in front of the lab tests. For example, DeValois and DeValois [21] suggested that stronger assimilation should be found for color in comparison with grayscale stimuli, because of the lack of lateral inhibition in chromatic receptive fields. However, the results by de Weert and Spillmann [22] showed that it was not the case.

Another approach that is not necessarily opposite to the bottom-up interpretation suggests that assimilation is primarily generated by more central mechanisms of visual processing, such as figure-ground segmentation [23, 24] and observer expertise [25]. It is also important to note that assimilation received interest based on the White illusion [26]. It has been suggested that assimilation depends on the existence of T-junctions that produce a perception of figure-ground segregation [27, 28]. T-junctions seem, for example, to affect also illusory motion [29]. However, this explanation was not supported by the lab test, which demonstrated that assimilation effects can also be seen in versions of White's display where T-junctions have been completely removed [30]. More recently, Soranzo et al. [31] supported the central mechanism explanation for assimilation by testing that in stroboscopic conditions. In 2010 Rude [32] proposed an intriguing computational neural model that includes the effect of the top-down attentional control in explaining the assimilation effect.

In summary, assimilation is an interesting phenomenon for which a convincing explanation that can keep all the experimental results under the same theoretical umbrella is still missing. However, several important steps were taken for the purpose of explaining how the brain interprets these patterns.

Cross-References

- [Afterimage](#)
- [Al-Haytham \(Alhazen\), Abū 'Alī al-Ḥasan ibn al-Ḥasan ibn](#)
- [Anchoring Theory of Lightness](#)
- [Color Contrast](#)
- [Complementary Colors](#)
- [Simultaneous Color Contrast](#)

References

1. Gori, S., Giora, E., Stubbs, D.A.: Perceptual compromise between apparent and veridical motion indices: the Unchained-Dots illusion. *Perception*. **39**, 863–866 (2010)
2. Gori, S., Giora, E., Yazdanbakhsh, A., Mingolla, E.: A new motion illusion based on competition between two kinds of motion processing units: the Accordion-Grating. *Neural Netw.* **24**, 1082–1092 (2011)
3. Kitaoka, A., Ashida, J.: Phenomenal characteristics of the peripheral drift illusion. *Vision*. **15**, 261–262 (2003)
4. Gori, S., Giora, E., Pedersini, R.: Perceptual compromise between apparent and veridical motion indices: the Unchained-Dots illusion. *Acta Psychol.* **129**, 399–409 (2008)
5. Gregory, R.L.: Knowledge in perception and illusion. *Phil. Trans. R. Soc. London*. **352**, 1121–1127 (1997)
6. Eagleman, D.M.: Visual illusions and neurobiology. *Nat. Rev. Neurosci.* **2**, 920–926 (2001)
7. Spillmann, L.: Phenomenology and neurophysiological correlations: two approaches to perception research. *Vis. Res.* **49**, 1507–1521 (2009)
8. Gori, S., Agrillo, C., Dadda, M., Bisazza, A.: Do fish perceive illusory motion? *Sci. Rep.* **4**, 6443 (2014). <https://doi.org/10.1038/srep06443>
9. Agrillo, C., Gori, S., Beran, M.J.: Do rhesus monkeys (*Macaca mulatta*) perceive illusory motion? *Anim. Cogn.* **18**(4), 895–910 (2015)
10. Kingdom, F.A.A.: Simultaneous contrast: the legacies of Hering and Helmholtz. *Perception*. **26**, 673–677 (1997)
11. Agostini, T., Galmonte, A.A.: New effect of luminance gradient on achromatic simultaneous contrast. *Psychon. Bull. Rev.* **9**, 264–269 (2002)
12. Gori, S., Stubbs, D.A.: A new set of illusions – the dynamic luminance gradient illusion and the breathing light illusion. *Perception*. **35**, 1573–1577 (2006)
13. Anstis, S., Gori, S., Wehrhahn, C.: Afterimages and the breathing light illusion. *Perception*. **36**, 791–794 (2007)
14. Gori, S., Giora, E., Agostini, T.: Measuring the breathing light illusion by means of induced simultaneous contrast. *Perception*. **39**, 5–12 (2010)

15. von Bezold, W.: Die farblehrer im Hinblick auf Kunst und Kuntsgewerbe. Westermann, Brunswick (1874)
16. Giora, E., Gori, S.: The perceptual expansion of a filled area depends on textural characteristics. *Vis. Res.* **50**, 2466–2475 (2010)
17. Vannucci, M., Gori, S., Kojima, H.: The spatial frequencies influence the aesthetic judgment of buildings transculturally. *Cogn. Neurosci.* **5**(3–4), 143–149 (2015). <https://doi.org/10.1080/17588928.2014.976188>
18. Yazdanbakhsh, A., Gori, S.: A new psychophysical estimation of the receptive field size. *Neurosci. Lett.* **438**, 246–251 (2008)
19. Helson, H.: *Adaptation-Level Theory*. Harper & Row, New York (1964)
20. Reid, R.C., Shapley, R.: Brightness induction by local contrast and the spatial dependence of assimilation. *Vis. Res.* **28**, 115–132 (1988)
21. DeValois, R.L., DeValois, K.K.: *Spatial Vision*. Oxford University Press, Oxford (1988)
22. de Weert, C.M.M., Spillmann, L.: Assimilation: asymmetry between brightness and darkness? *Vis. Res.* **50**, 2466–2475 (1995)
23. Musatti, C.: Forma e assimilazione. *Archivio Italiano di Psicologia*. **9**, 213–269 (1931)
24. de Weert, C.C.M., van Kruysbergen, N.: Assimilation: central and peripheral effects. *Perception*. **26**, 1217–1224 (1997)
25. Kanizsa, G.: *Organization in Vision: Essays on Gestalt Perception*. Praeger, New York (1979)
26. White, M.: A new effect of pattern on perceived lightness. *Perception*. **8**, 413–416 (1979)
27. Anderson, B.L.: A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. *Perception*. **26**, 419–453 (1997)
28. Todorović, D.: Lightness and junctions. *Perception*. **26**, 379–394 (1997)
29. Gori, S., Hamburger, K., Spillmann, L.: Reversal of apparent rotation in the Enigma-figure with and without motion adaptation and the effect of T-junctions. *Vis. Res.* **46**, 3267–3273 (2006)
30. Yazdanbakhsh, A., Arabzadeh, E., Babadi, B., Fazl, A.: Munker–White-like illusions without T-junctions. *Perception*. **31**, 711–715 (2002)
31. Soranzo, A., Galmonte, A., Agostini, T.: von Bezold assimilation effect reverses in stereoscopic conditions. *Perception*. **39**, 592–605 (2010)
32. Rudd, M.E.: How attention and contrast gain control interact to regulate lightness contrast and assimilation: a computational neural model. *J. Vis.* **10**, 1–37 (2010)

Automotive Front Lighting System (Status 2020)

T. Q. Khanh and J. Kobbert

Laboratory of Lighting Technology, Technische Universität Darmstadt, Darmstadt, Germany

Synonyms

Advanced front lighting systems (AFS) – intelligent light; Bend light – cornering light; Glare-free high beam – adaptive driving beam (ADB); Matrix beam; Multi-beam; Pixel light; Segmented high beam

Definition

Automotive lighting in general includes all lighting functions installed on motor vehicles. While this includes rear and signaling devices, this chapter will focus only on front lighting systems. The main goal for automotive headlamps is to increase safety for drivers at night without glaring other traffic participants. For this reason, automotive manufacturers have constantly developed new headlamp technologies. Beginning with the historically and currently still relevant light sources for automotive front lighting systems such as tungsten halogen, HID, and LED light sources, different steps in the evolution of headlamp systems and their improvements are discussed in this paper. Going away from light sources, the historical evolution of intelligent headlamps and their light functions are discussed. Starting with the now widely available static and dynamic bend light, this chapter will work its way up to the latest headlight function – the road projection that is supposed to be market ready in 2020.

Aura

► Scintillating Scotoma

Introduction

Over the last years, the traffic density has increased significantly. With more and more vehicles driving on our roads simultaneously, one of the challenges for

car manufacturers is to ensure that traveling by car is as safe as possible. While this has led to improvements in the overall vehicle safety with additional active (e.g., emergency braking assistance) and passive safety features like improved crash stability, one key feature for safe driving at night is improved headlamps. Since the light situations between daytime driving and nighttime driving are completely different, during the day a global illumination from the sun generally lights the whole scene in front of the driver, enabling him to perceive all the necessary information for driving. At night the driver has to rely on artificial illumination. While in cities road illumination is present and allows for nearly the same overall light situation as during the day, on country roads or motorways, the driver has to rely on the headlamps of his vehicle. In order to make driving in those situations as safe as possible, headlamp manufacturers and car OEMs have developed and introduced new lighting technologies. The following sections will therefore introduce these new technologies. These sections will be split thematically into light sources and light functions.

Light Sources

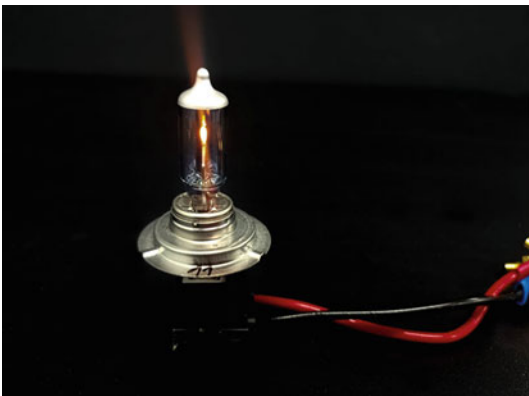
In order to deliver the necessary amount of light for the driver to safely detect obstacles on the road, several new light sources were introduced to the mass market over the last decades. This

section will chronologically introduce those light sources, starting with the oldest and still most common one.

Tungsten Halogen Lamps

The oldest currently available light source for automotive headlamps is a tungsten halogen headlamp. Since this light source is the cheapest to produce, up until today it still holds the largest market share in automotive headlamps with over 80%. Figure 1 shows a standard H7 tungsten halogen light source for automotive headlamps on the left side and a luminance image of the same H7 light source on the right side. For better image quality, the tungsten halogen light source has been dimmed.

In a tungsten halogen headlamp, a small spiral-wound filament is used for light emission. Applying 12 V, as standard in automotive use, this filament is heated up to approximately 3000–3200 K which also represents the correlated color temperature (CCT) and starts to glow. Due to the thermal material boundaries and global regulations regarding tungsten halogen headlamps, their total luminous flux, the total light they emit, is limited to 1500 lumens. However, due to the nature of a so-called Planckian thermal radiator, most of the radiation it emits is not visible to the human eye, leading to a very low efficiency of only 20–25 lumens per watt.



Automotive Front Lighting System (Status 2020), Fig. 1 Left: H7 tungsten halogen light source. Right: Luminance image of the same H7 light source

High-Intensity Discharge (HID)

In order to increase the total luminous flux and the efficiency, high-intensity discharge (HID) light sources and headlamps were introduced in 1991. Compared to the filament in the tungsten halogen light sources, HID light sources use gas discharges to produce light. Since this requires a very high voltage, the 12 V from the car was not enough anymore, and a ballast is required in order to achieve about 10 kV. Nevertheless, HID headlamps have several advantages compared to tungsten halogen light sources, with the most important one being a higher luminous flux of up to 3200 lumens, a higher color temperature of around 4200–4500 K, and therefore an increased brightness perception for the driver and a much higher efficiency of up to 90 lumens per watt. The increase in total luminous flux also led to the introduction of two safety features in those HID headlamps that prevent other traffic participants from being glared. All headlamps, with light sources that produce more than 2000 lumens, require a headlamp cleaning device and an automated leveling system. The headlamp cleaning device avoids the possibility of dirt on the outer headlamp lens from scattering light into regions, where other traffic participants might be. The automated leveling system measures any vehicle inclination due to increased load in the trunk, for example, and automatically lowers the headlamp's aim.

However, since those two measures cost both weight and room in the vehicle, things car manufacturers never want to give up, 25-watt HID headlamps with a luminous flux of just below 2000 lumens, were introduced. Nevertheless, HID headlamps were soon replaced by their

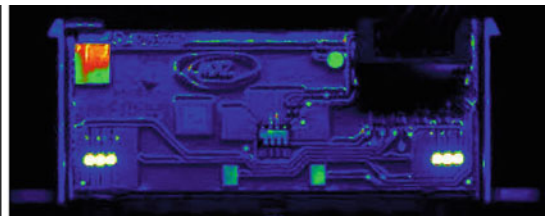
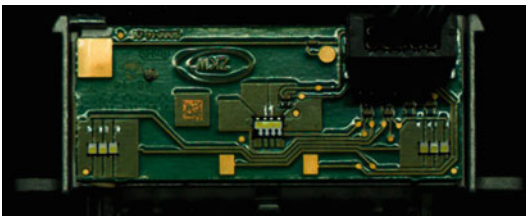
successor and are currently not being used in new car generations anymore.

Light-Emitting Diodes (LEDs)

While light-emitting diodes (LEDs) are being used in general lighting since the early 2000s, it took until 2007 for the first headlamp to use LEDs for passing beam (Lexus LS 600h and Audi R8) and until 2010 (Audi A7, Mercedes Benz CLS) for a headlamp to solely utilize LEDs for all light functions like passing beam, driving beam, day-time running light, and turn indicator as well as several advanced lighting functions that will be explained in the next sections.

Compared to the thermal radiators (tungsten halogen lamps) and gas discharge lamps (HID), LEDs represent the third type of light source for automotive headlamps – the solid-state light source. The principle of solid-state lighting will not be explained here since this would go beyond the scope of this chapter but can be found in Ref. [1].

One of the main challenges when using LEDs as a main light source is the relatively low luminous flux of only around 1000 lumens – only a third of what an HID headlamp can produce. Due to this limitation, modern headlamps use several single LED chips for their light functions. As an example, the light module for low and high beam is shown in Fig. 2 with a photo of the layout on the left side and a luminance image on the right side. For measurement reasons, the LEDs in the luminance image on the right side are dimmed down to 1%, and only the passing beam LEDs are switched on. The driving beam LEDs are in the center of the module.



Automotive Front Lighting System (Status 2020), Fig. 2 LED module for a modern headlamp with ten LED chips for passing (2×3 LEDs) and driving beam (1×4 LEDs)

The main advantage of LEDs however is the high efficiency of at around 80–100 watts per lumen (automotive grade LEDs); the long-term reliability of around 10,000 h, thus exceeding the lifetime of a vehicle which drives approximately around 6,000 h (300,000 km at an average speed of 50 km/h); and the high flexibility and the low required volume.

Light Amplification by Stimulated Emission of Radiation (Laser)

To further increase the viewing distance for drivers, car manufacturers such as BMW, Audi, and Jaguar Land Rover started using light amplification by stimulated emission of radiation (laser) diodes as additional light sources for the driving beam. Since the usage of a laser light source in a vehicle is accompanied by a high amount of safety regulations and a very complex setup, this is only used in addition to a base LED driving beam. The advantage of a laser light source is that it can reach much higher luminance values compared to LED, HID, and tungsten halogen headlamps since the light-emitting area is comparably small. This is then used to produce a small hot spot in the center of the driving beam to allow drivers a larger viewing distance.

Compared to LED high beam, this can lead to an increase of about 60% as indicated by Fig. 3 where the detection distances of a pedestrian like dummy next to the road are measured [3].

Due the high cost and effort involved in using a laser light source, these headlamps are currently only found in a very select number of vehicles, and this will therefore not be discussed any further.

Advanced Front Lighting Technologies

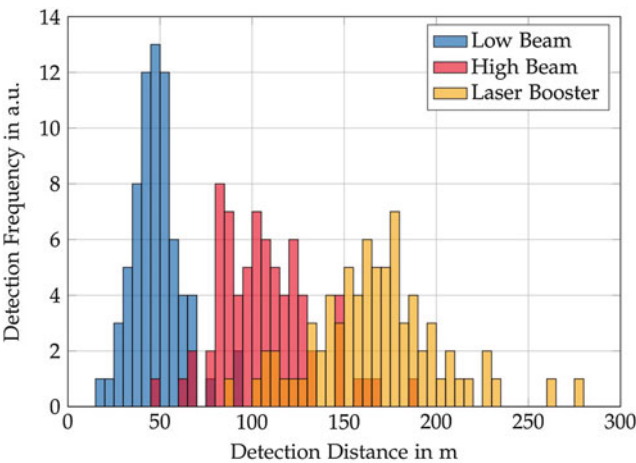
Similar to the advancements in the light sources, new advanced front lighting functions have been implemented and introduced to the public over the last two decades. Figure 4 shows a brief overview over the currently available headlamp features.

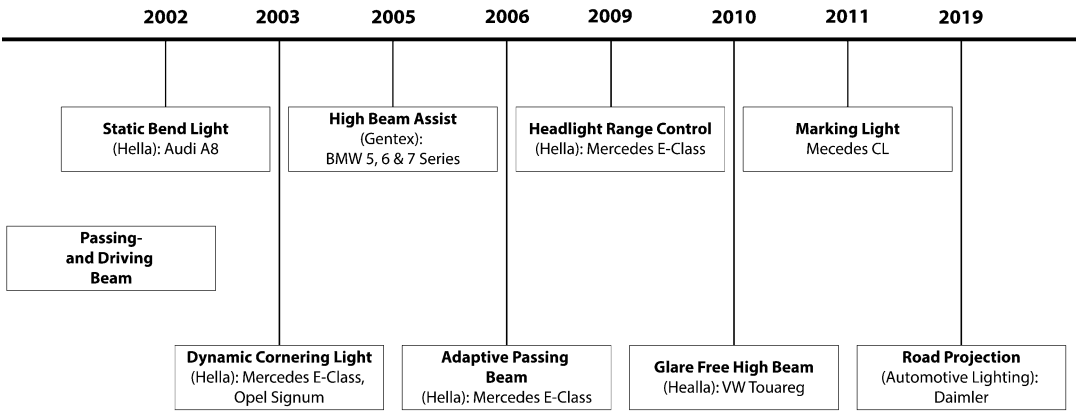
This section will then explain in some detail each lighting function and the resulting benefit from it.

Static Bend Light/Cornering Light

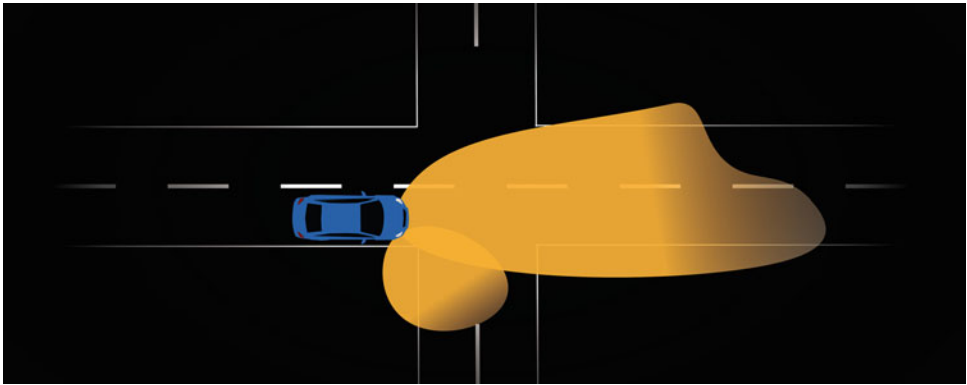
In 2002 the first additional light function that was introduced was the static bend light or cornering light. Usually this is implemented by an additional light module that illuminates very wide angles as shown in Fig. 5. The idea for this is that when turning into a junction, the usual light distribution will not illuminate the sides sufficiently in order to detect pedestrians or cyclists that are crossing the road. The static bend light will then activate at very low speeds (<30 km/h) depending on the steering angle thus illuminating the intersection better.

Automotive Front Lighting System (Status 2020), Fig. 3 Detection distances for a pedestrian next to the road for LED low beam (blue), LED high beam (red), and high beam with added laser module (yellow) [2]





Automotive Front Lighting System (Status 2020), Fig. 4 Chronological introduction of adaptive front lighting technologies since 2002 up to today



Automotive Front Lighting System (Status 2020), Fig. 5 Low beam and static cornering light

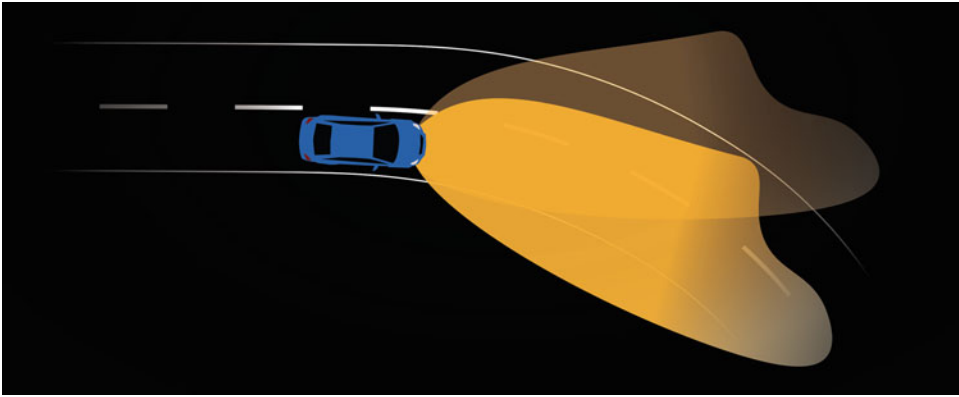
Dynamic Cornering Light

Only 1 year later, the dynamic cornering light was introduced. The issue with the “normal” passing and driving beam light distributions when driving through corners is that the illumination will always be tangential to the actual corner, making it difficult to detect objects on the outside of the corner. To solve this issue, Daimler, Hella, and Opel introduced the dynamic bend light. For this light function, the low beam module is able to be swiveled by up to about 15° using stepper motors, thereby illuminating the bend better. This is illustrated in Fig. 6, where a static low beam is shown in dark orange in the background and a dynamic bend light is shown in the foreground.

This motion is dependent on the speed the vehicle is going, the steering angle, and the implementation by the manufacturer. While this does solve some of the issue, one still existing problem is that the swiveling will only start once the vehicle is in the corner. Approaches where GPS signals were used to predict bends failed due to the inaccuracy of such systems, and the automotive industry is still working on better solutions.

High Beam Assist

One issue that was shown by Sprute [4] is that human drivers tend to underuse the driving beam. His study shows that driving beam was only used 29.5% of the drive’s duration, while the recorded



A

Automotive Front Lighting System (Status 2020), Fig. 6 Static passing beam (dark light distribution in the back) and dynamic bend light (foreground)

maximum potential for high beam driving is at 60.5%, therefore more than double. Since passing beam should only be used during nighttime when passing another vehicle or while driving behind another vehicle since the viewing distance is limited, this results in a severe safety issue.

A first solution for this issue was introduced by BMW in 2005 with the high beam assist. A vehicle equipped with this feature automatically uses the driving beam when possible. For this, a camera is placed behind the windshield, usually inside or right behind the rearview mirror, which checks the road ahead for oncoming or preceding traffic. Due to the camera-based nature of this system, humans are still better in judging, when a vehicle is approaching and could, especially in cornering situations, often react earlier than this system.

Adaptive Passing Beam

All previous systems are based around one standard passing beam light distribution. However, it is easy to imagine that this singular light distribution cannot be the optimal light distribution in all situations. For that, Hella and Daimler introduced the adaptive passing beam. Depending on the situation, a different light distribution will be loaded by the vehicle. Exemplary light distributions are:

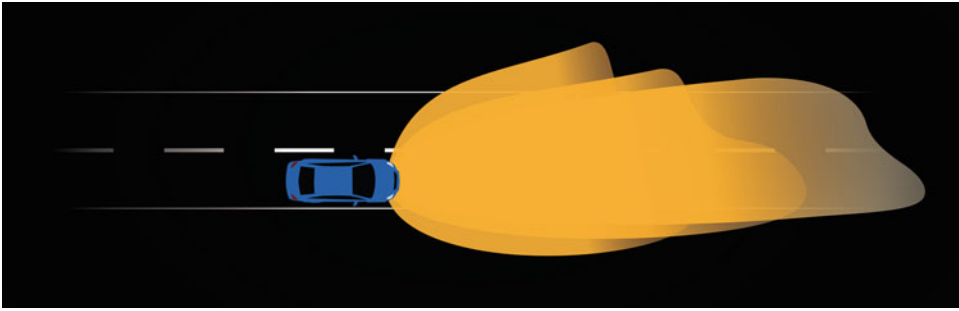
- Motorway light: the cutoff line of the low beam is moved up to the horizon by 0.25° :

- The higher driven speed requires more viewing distance. At the same time, fore field illumination is not required anymore since the driver needs to focus on the far field. To activate this function, the vehicle needs to be on a driving more than 100 km/h.
- Bad weather light: the low beam is lowered and swiveled to the side of the road:
 - Due to the higher reflectance of the road when wet, oncoming traffic could perceive more glare. At the same time, the darker road makes it more difficult for the drivers to orientate themselves on the road. To counter that, lowering the passing beam and swiveling it to the side increases the illumination of objects next to the road, thereby enabling the drivers to orientate themselves there. The intensity of the headlamp on the inner road side may be reduced as well to lower the glare potential for oncoming traffic.

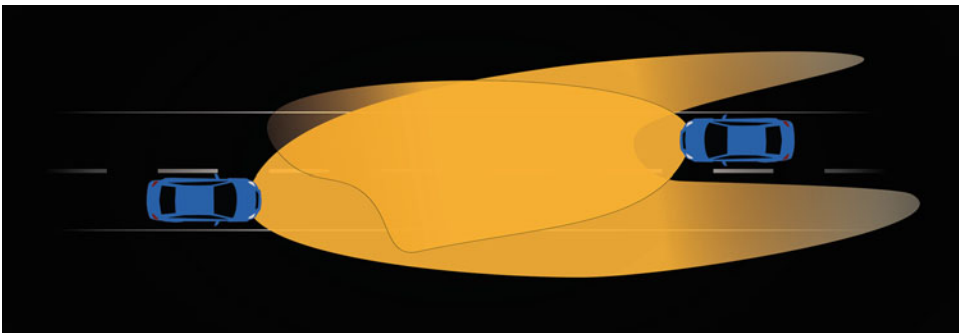
Examples of these light distributions are shown in Fig. 7.

Headlight Range Control

While the highway light as introduced in the section above already bridges the gap between passing and driving beam slightly, this is still a digital selection between three levels of viewing distance. To achieve even better results, a headlight



Automotive Front Lighting System (Status 2020), Fig. 7 Adaptive passing beam light distributions. Bad weather light, normal passing beam, and motorway light



Automotive Front Lighting System (Status 2020), Fig. 8 Illustrated glare-free high beam light distribution (left) for an oncoming vehicle with passing beam (right)

range control was proposed in 2009. Again a camera detects, if other vehicles are on the road in front of the vehicle, and now also finds the exact position of head- or taillights. A stepper motor is now able to put the cutoff line of the low beam directly under the position of the head- and taillights of the approaching or preceding vehicle.

Glare-Free High Beam

The next and current step in evolution focusing on the viewing distance of the driver by optimizing the high beam usage is the so-called glare-free high beam. This system, introduced in 2010 by Volkswagen, enables the driver to use high beam at nearly all times when driving on country roads or motorways. Again a camera detects if and where other vehicles are. Instead of switching from driving beam to passing beam now, a shadow is inserted into the driving beam light

distribution to mask out other traffic participants. As illustrated in Fig. 8, this leads to light being sent to the sides of oncoming or preceding traffic, enabling the driver to much higher viewing distances.

The *THM, University of Applied Sciences, Gießen* has shown that the detection distances of objects beside the road with glare-free high beam can be increased by a mean of 32 m on average [5]. *Zydek* shows that for the detection of human-like figurines, the detection distance with passing beam is at 88.4 m, at 141.3 m with driving beam, and at 120.4 m with glare-free high beam, thereby closing the gap between the two basic light functions. *Zydek* also shows that the glare level perceived by oncoming traffic when driving with glare-free high beam is not significantly different from passing beam in any situation [12].

Technologically, this function can be implemented in two different ways. A mechanical solution uses a

sliding aperture to create a shadow in the driving beam light distribution. With the emerging of LEDs as light sources and their previously mentioned high flexibility, an electrical solution, where single LEDs are used to create single areas in the light distribution, was introduced as well. Simplified speaking, single LEDs are turned off to create the necessary shadow in the light distribution. LED-based glare-free high beam solutions are available with different number of LED segments, starting as low as 4 segments per headlamp and currently going as high as 82 segments per headlamp. However, different companies have already shown prototype solutions with thousands of segments per headlamp [6].

Both solutions come with advantages as well as with disadvantages. The mechanical solution is only capable of creating a single shadow. However, by using a modular approach in both headlamps, the size of this shadow is flexible. Using single LEDs, on the other hand, enables the car manufacturer to create multiple shadows; however, the size of these shadows is always limited by the size of the segment each LED produces. Another advantage of the electrical version over the mechanical one is that no moving parts are required thus increasing the reliability.

Marking Light

All new light functions so far focused on bridging the gap between passing and driving beam. A different approach was presented by BMW in 2011. Instead of delivering as much light as possible, the approach here is to use a small concentrated light spot to highlight important areas. In order to do so, an infrared camera is used to detect

pedestrians or animals beside or on the road at distances where the human eye is not able to do so. A small LED spot is then switched on for illuminating the object. In order to illuminate the target, this light spot is moveable on at least one axis by stepper motor. Schneider shows that this approach actually leads to better reaction times when detecting objects compared to driving beam [7]. However, this approach reintroduces moveable parts and requires additional hardware such as a night vision camera.

Road Projection

The latest addition to intelligent headlamp features is “road projections” where symbols or markings are projected into the low beam area. This feature is used to provide the driver with an alternative to a head-up display by providing simple information or warning signs directly onto the road as an additional way of conveying information to the driver similar to a head-up display. Additional features include the possibility to mark the width of the vehicle when driving through narrow road sections such as road works.

In order to do so, a digital mirror device (DMD) by Texas Instruments, known for its use in home theater projectors, which is able to produce millions of individually addressable pixels, is set into the headlamp as well and illuminated with high-power LEDs. Similar to the technology in conventional projectors, small mirrors reflect the light either onto the road or into a “light trap.” Controlling these mirrors, different symbols can be displayed on the road as shown in Fig. 9 [9].



Automotive Front Lighting System (Status 2020), Fig. 9 Symbol projection as presented by Daimler in their digital light concept [8]. (© 2019 Daimler AG. All rights reserved)

While one concern with this technology was that other road users might be distracted or even misinterpret these projections as signals for them, studies have shown that the projections are only visible from the driver's position and other road users don't even perceive them, let alone think they are addressed at them [10, 11].

Cross-References

- [Autonomous Vehicle \(AV\) Lighting Systems](#)
- [Electrical Control Gear for Lamps](#)
- [Glare](#)
- [Goniophotometer](#)
- [Halogen Lamp](#)
- [Illuminance Meter](#)
- [Instrument: Photometer](#)
- [Light Distribution](#)
- [Light-Emitting Diode, LED](#)
- [Luminance Meter](#)
- [Photodetector](#)
- [Road Lighting](#)
- [Standard Measurement Geometries](#)
- [Vision and Age](#)
- [Xenon Lamp](#)

References

1. Khanh, T.Q., Bodrogi, P., Vinh, T.Q., Winkler, H.: LED Lighting: Technology and Perception. Wiley, Weinheim. ISBN: 978-3527412129 (2014)
2. Kobbert, J.: Optimization of automotive light distributions for different real life traffic situations. Dissertation, Technische Universität, Darmstadt. Online edition: <https://tuprints.ulb.tu-darmstadt.de/8382> (2019)
3. Kobbert, J., Kosmas, K., Polin, D., Englisch, D., Schneider, K., Khanh, T.Q., Society of Automotive Engineers (SAE) (Urheber): Field test of visibility distances and recognition rates – comparison of LED and Laser Systems. In: SAE Technical Papers. Proceedings of the SAE World Congress 2016, Detroit, 12–14 Apr 2016
4. Sprute, J.H.: Entwicklung lichttechnischer Kriterien zur Blendungsminimierung von adaptiven Fernlichtsystemen. Dissertation, Technische Universität, Darmstadt (2012)
5. THM Gießen & LSS: Prove benefit of ADB systems under real life conditions. In: Technical Report. Available (online): <https://clepa.eu/wp-content/uploads/2016/04/Benefits-of-Adaptive-Driving-Beam-systems-under-eal-life-conditions-1.pdf> (2015)
6. Trommer, J., et al.: New possibilities with μ AFS modules – the path to high-resolution full-matrix headlamps. In: 12th International Symposium on Automotive Lighting (ISAL 2017), pp. 333–341 (2017)
7. Schneider, D.: Markierungslicht – eine Scheinwerferlichtverteilung zur Aufmerksamkeitssteuerung und Wahrnehmungssteigerung von Fahrzeugführern. Herbert Utz Verlag, München (2011)
8. Daimler: World premiere in the Mercedes-Maybach S-Class: DIGITAL LIGHT: the light of the future hits the road online: <https://media.daimler.com/marsMediaSite/en/instance/ko/World-premiere-in-the-Mercedes-Maybach-S-Class-DIGITALLIGHT-the-light-of-the-future-hits-the-road.xhtml?oid=33898368>. Accessed 12 Mar 2020
9. Rosenhahn, E.O., et al.: Digital light millions of pixels on the road. ATZ Worldwide. **120**(11), 42–45 (2018)
10. Budanov, M., Neumann, C.: Road projections as a new and intuitively understandable human-machine interface. Adv. Opt. Technol. **8**(1), 77–84 (2019)
11. Rosenhahn, E.-O., Link, F.: Traffic safety benefits provided by high resolution headlamp system. In: ISAL Symposium 2019, Darmstadt (2019)
12. Zydek, B.: Blendungsbewertung von Kraftfahrzeugscheinwerfern unter dynamischen Bedingungen. Dissertation, Technische Universität, Darmstadt (2014)

Autonomous Vehicle (AV) Lighting Systems

J. Kobbert and T. Q. Khanh
Laboratory of Lighting Technology, Technische Universität Darmstadt, Darmstadt, Germany

Synonyms

[Autonomous vehicles – automated vehicles](#); [AV](#); [Highly automated vehicles](#)

Definition

Lighting systems for autonomous vehicles have completely different requirements when compared to the traditional automotive lighting systems of today. This chapter will focus on the changes necessary for vehicle lighting with the

introduction of autonomous or highly automated vehicles. While the current lighting systems for daily use are intended to enable the driver to safely steer the vehicle at night, an autonomous vehicle does not have a driver that needs to be able to view the road ahead. However, the necessity for lighting systems is still given since autonomous vehicles heavily rely on cameras that need light to function properly. Additionally, new studies show that the majority of people would like some kind of communication between autonomous vehicles and other human road users. Realizing these communication devices with light is one of the key challenges for the future.

Introduction

Over the last couple of years, more and more vehicles with highly automated functions have been introduced to the market. Starting with lane assist up to automated driving on certain types of road, the trend clearly shows that within the next years, fully automated vehicles will be part of our traffic environment.

While automotive lighting up until now focused mainly on the task of delivering light for the driver, to ensure safe driving, as well as a rudimentary signaling to other traffic participants with braking lights and turn indicators, lighting for automated vehicles will not have “drivers” as known today, and therefore the focus on lighting will shift. Two major areas are already in development by car manufacturers as of today. First of all, signaling has one of the highest priorities. With different approaches, car manufacturers are trying to communicate with other traffic participants in order to minimize misunderstanding, increase traffic flow, and even reduce the chance of accidents involving automated vehicles. The second large area of interest is lighting for the passengers. With no driving task, riding in an automated vehicle opens up the possibility for the passengers to pursue other tasks, ranging from relaxing or even sleeping to a mobile office. Developing smart interior lighting that suits those cases and puts the passenger and his task in the focus is the second big challenge for automotive

original equipment manufacturers (OEMs). Furthermore, all automated vehicles rely on several sensors, from cameras to RADAR and LIDAR sensors. Optimizing front lighting for those sensors to perform as good as possible even during the night is another big challenge when talking about lighting systems for automated vehicles.

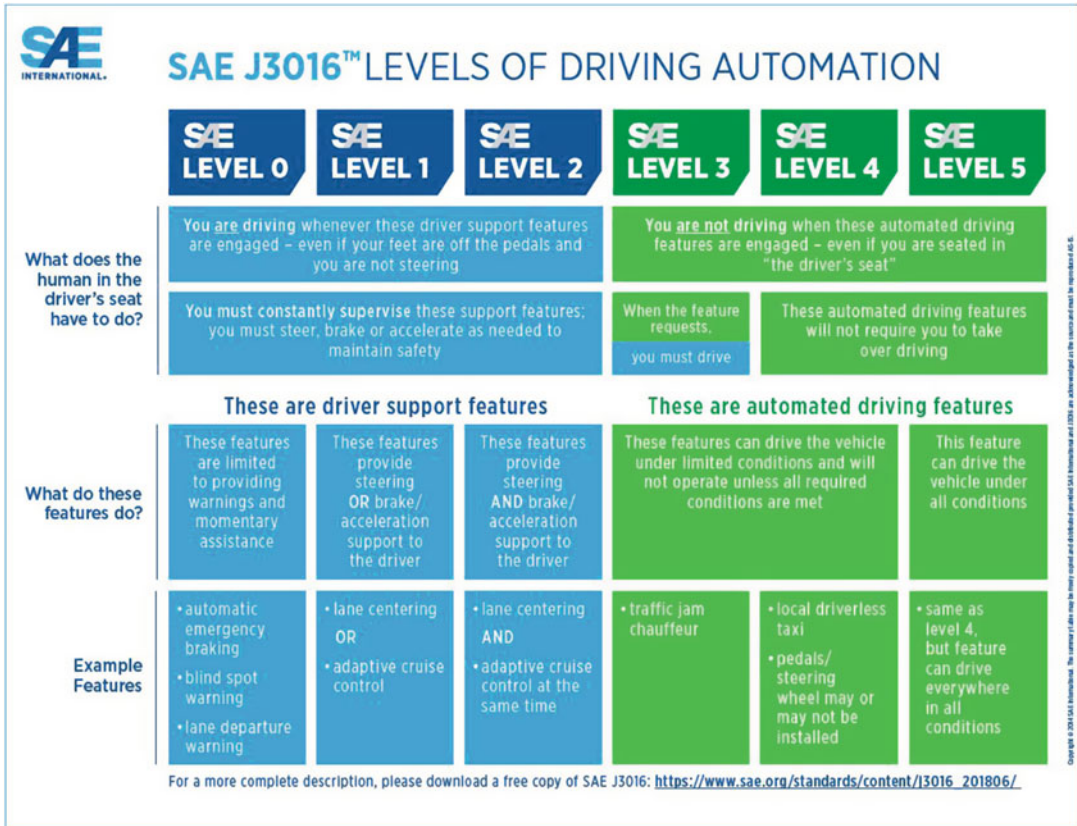
This chapter will start by explaining the different levels of highly automated vehicles and the issues arising with highly automated vehicles driving in mixed traffic with other human road users (HRUs). The next section will then put emphasis on the different possibilities of using light for communication with other HRUs. The third section will explain the current development of interior human centric lighting for automated vehicles before the last section focuses on optimizing light for sensors.

Lighting for Highly Automated Vehicles

In order to discuss the different needs for lighting on automated vehicles, the first section will shortly introduce the different types of automated vehicles and motivate the need for communication devices on these automated vehicles.

Different Levels of Automated Vehicles

The Society of Automotive Engineers International (SAE International), a US-based standard developing association, has defined different levels of automated driving ranging from level 0 to level 5 (see SAE J3016 for reference). Level 0 describes vehicles with no assistant system. The driver performs all dynamic driving tasks (DDTs). Level 1 vehicles incorporate driver assistance features; level 2 vehicles have partial-automated driving features. Only level 3 and above vehicles can perform full DDTs. Level 3 vehicles can drive automated under certain conditions but still require a user to be able to take over; level 4 vehicles can perform automated driving under some conditions but do not require a driver to take over under these situations. Level 5 vehicles are vehicles with full driving automation. All these levels are summarized in Fig. 1.



Autonomous Vehicle (AV) Lighting Systems, Fig. 1 Different levels of automated vehicles according to SAE J3016. (©2019 SAE International. All rights reserved)

When referring to highly automated vehicles from now on, level 4 or level 5 vehicles are meant.

Reviewing exemplary traffic encounters, like a pedestrian crossing a road in front of a vehicle, it becomes obvious that some sort of communication between the driver of the vehicle and the pedestrian, while not absolutely necessary, is certainly helpful for the pedestrian.

Now please imagining a newly introduced automated vehicle without a human driver in the same situation. While the pedestrian looks to the approaching vehicle, the passenger may not be paying attention or even be in the vehicle. While this might not lead directly to a dangerous situation, this might leave the pedestrian wondering and confused about the next actions of the automated vehicle.

This combined with the increase in reported accidents involving automated vehicles over the last years leads to a decrease in trust into automated vehicles [1]. Openly communicating in an easy to understand way what the next intentions of the vehicle are could help increase the trust and acceptance toward automated vehicles [2].

Communication Strategies for Automated Vehicles with Human Road Users

In this section, the different possible strategies for light-based communication between highly automated vehicles (AVs) and HRUs are being explained. For this, the communication approach is split into different groups. The first section is dedicated to display-based communication where symbols, numbers, or even text is displayed on the



Autonomous Vehicle (AV) Lighting Systems, Fig. 2 Display integrated into a highly automated vehicle by Daimler [3]. (© 2019 Daimler AG. All rights reserved)

outside of the vehicle. The second section discusses light stripes before the third section introduces road projection as the last means to send messages from the AV to the HRU.

Display-Based Communication

When thinking about communicating between an AV and other road users, the most obvious idea is to use some sort of display. While this might not necessarily mean a high-resolution LCD or OLED panel that is wrapped around a vehicle, prototypes using LED matrices as means of integrating and displaying complex messages have been presented to the public by different companies. Figure 2 shows an example by Daimler of integrating a small display into the area of the grill for an AV.

While written text as shown in the example by Daimler is only feasible at short distances, larger, easier to understand symbols similar to traffic signs might be used for larger distances as well. However, these symbols and texts need to be universally understandable and intuitively designed.

Light Stripe-Based Communication

A different approach, shown, for example, by Ford and Nissan [4, 5], is the possibility to use simplified signals by implementing different light stripes around the vehicle. This is shown in Fig. 3 where a light stripe is implemented around the test vehicle.

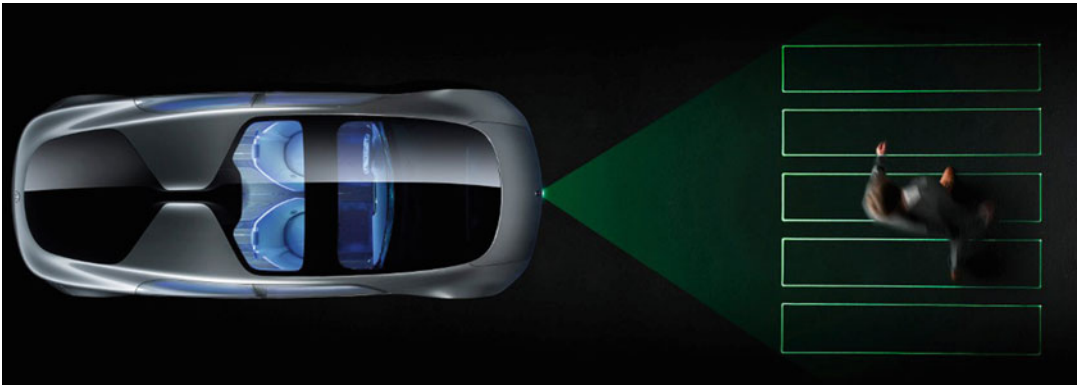
By using different colors or animations on these strips, different messages such as “driving autonomously” and “I have seen you” or driving intentions “waiting,” “accelerating,” and “stopping” could be displayed. These animations might include blinking, swiping, or “loading bars.”

Projection-Based Communication

While both display- and light stripe-based communications require the HRU to look directly at the vehicle in order to perceive the message from the AV, projection-based communication provides the opportunity to send information in front of the HRU, thereby enabling communication even with distracted HRUs. In contrast to that, different environmental conditions, such as road surfaces, can have a major influence on the projection visibility. The technological basis for this has already been shown by Daimler in their Maybach’s S-Class. As described in the chapter ▶ “Automotive Front Lighting System (Status 2020)”, a DMD-based imaging device can be placed inside a headlamp to allow for detailed (megapixel) images to be projected onto the road. While the section in ▶ “Automotive Front Lighting System (Status 2020)” describes the use-case similar to the one of a head-up display, highly automated vehicles could use the same technology to project information for other road users onto the road. An example for this is again provided by Daimler in the form of the F 015 prototype [6]. As shown in

Autonomous Vehicle (AV) Lighting Systems,

Fig. 3 Test vehicle by Ford with LED stripes on the roof that can show different animations in order to test pedestrian behavior when encountering an AV [4]. (© 2018 The Ford Motor Company)



Autonomous Vehicle (AV) Lighting Systems,
Fig. 4 Highly automated vehicle prototype by Daimler, the F 015 projecting a crosswalk onto the road as a sign that

crossing in front of the vehicle is safe [7]. (© 2019 Daimler AG. All rights reserved)

Fig. 4, the vehicle is capable of projecting a crosswalk for a pedestrian as a sign for the HRU that crossing the road in front of the vehicle is safe.

Light for Sensors

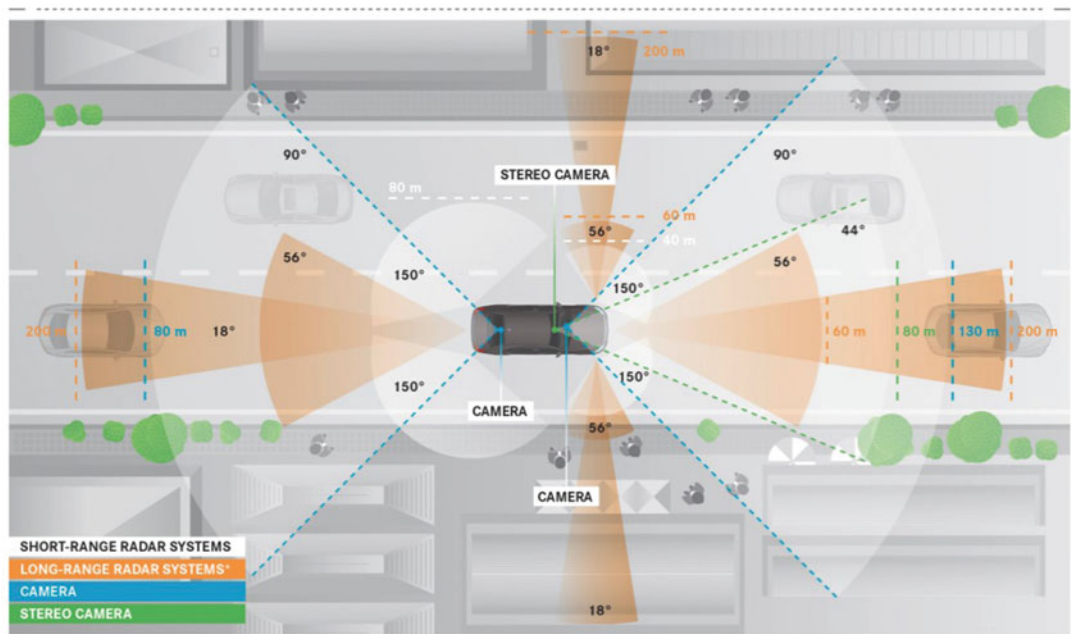
As mentioned in the introduction and as described in more detail in the chapter ► [“Automotive Front Lighting System \(Status 2020\)”](#), front lighting so far has been focused on increasing the detection distances of drivers while at the same time minimizing glare for other traffic participants. With the introduction of level 4 and level 5 vehicles, the vehicle can operate without any driver in them. In order to do that safely, a large sensor array is used

in order to measure the surroundings of the vehicle at all times. As an example the sensors of the Mercedes-Benz Intelligent Drive “Bertha” are shown in Fig. 5.

As shown by Daimler, a large area of the traffic space is covered by cameras either in single camera use, indicated in blue, or by stereo camera systems, indicated in green. Only small areas but large distances are covered by radar. While radar is not influenced by the surrounding light conditions, cameras are highly dependent on other light sources.

Using headlights and optimizing their light distribution to increase the performance of the

▲ **Networked sensor systems of the S 500 INTELLIGENT DRIVE research vehicle**
 Highly autonomous driving based on near-production sensor systems



Autonomous Vehicle (AV) Lighting Systems, Fig. 5 Sensors used in the Mercedes-Benz Intelligent Drive “Bertha” where cameras and radar sensors are used

for vehicle detection and mapping of the surroundings [8]. (© 2019 Daimler AG. All rights reserved)



Autonomous Vehicle (AV) Lighting Systems, Fig. 6 ESF prototype vehicle by Daimler (left) with the Daylight+ integration (right) [10]. (© 2019 Daimler AG. All rights reserved)

cameras now becomes one of the major developing points, and it has already been shown that using high-performance headlamps increases the confidence of object detection significantly [9].

Interior Lighting for Automated Vehicles

While the lighting systems discussed so far all focused on exterior lighting, new interior lighting

systems will emerge as well for AVs. Since none of the passengers needs to focus on driving and thereby detecting objects or keeping the vehicle safely on the road, optimizing the interior light on the new tasks that are being performed inside of the vehicle becomes a higher priority. While this sort of lighting has already been a major research point for home and office lighting [12], interior

lighting for vehicles has been scarcely researched up until now.

Daimler has presented prototypes and concepts that could be used in current vehicles as well. A high intensity light panel (Daylight+) is introduced in their concept vehicle ESF as part of the all-round driver safety (Fig. 6).

The goal by Daimler is to increase the brightness inside the otherwise comparably dark driver's cabin, thereby increasing the driver's alertness [10]. Audi increased the individualization for passengers by introducing "Matrix Interior Lighting Systems" to the market. Similar to the glare-free high beam systems discussed in the chapter ▶ "Automotive Front Lighting System (Status 2020)", individual LED segments are used for each passenger allowing them to customize the direction and the intensity of their individual lighting system [11].

However, compared to the state of the art in Human Centric Lighting (HCL) in home and office applications, these lighting systems are still at the beginning. In those applications light is intentionally used to increase the user's well-being, mood, and even health [11]. This can be done by simulating natural light – our body synchronizes with the natural sunlight, and sitting in dark offices with artificial light sources can get our body out of that natural rhythm which in turn can have negative effects on our health – or by finding the right lighting conditions for the task that is being performed at the moment. The right lighting condition can have a major effect on performance and productivity [11].

Further Remarks

While first automated vehicles are already on our roads, all previously discussed lighting systems are still in development. Especially using light as a means of communication between an AV and a HRU is still being investigated in numerous ways. The presented methods all include advantages and disadvantages that still need further research. Rasouli gives an overview about conducted studies and research methods [7]. Additionally to that, no regulatory basis for

these kinds of exterior lighting system has been released as of today since the research on what kind of signal in what kind of situation and by what kind of technology needs to be presented is not fully concluded.

Cross-References

- ▶ Automotive Front Lighting System (Status 2020)
- ▶ Electrical Control Gear for Lamps
- ▶ Glare
- ▶ Goniophotometer
- ▶ Illuminance Meter
- ▶ Instrument: Colorimeter
- ▶ Instrument: Photometer
- ▶ Instrument: Spectrophotometer
- ▶ Human-Centric Lighting
- ▶ Light-Emitting Diode, LED
- ▶ Luminance Meter
- ▶ Nonvisual Lighting Effects and Their Impact on Health and Well-Being
- ▶ Photodetector
- ▶ Road Lighting
- ▶ Standard Measurement Geometries
- ▶ Vision and Age

References

1. American Automobile Association: Vehicle technology survey – phase IIIB fact sheet (2018)
2. Singer, T.: Investigation and comparison of pedestrian behaviour in different encounter scenarios with automated vehicles. In: Proceedings of the 13th International Symposium on Automotive Lightning, Herbert Utz Verlag GmbH, München (2019)
3. Daimler: Autonomous concept car smart vision EQ fortwo: welcome to the future of car sharing. Available (online): <https://media.daimler.com/marsMediaSite/en/instance/ko/Autonomous-concept-car-smart-vision-EQ-fortwo-Welcome-to-the-future-of-car-sharing.xhtml?oid=29042725> (2017). Last Accessed on: 6 Jan 2020
4. Ford: Ford tests light-based visual language that could help autonomous vehicles communicate with pedestrians, Cologne. Available (online): <https://media.ford.com/content/fordmedia/feu/en/news/2019/02/06/ford-tests-light-based-visual-language-that-could-help-autonomou.html> (2019). Last Accessed on: 6 Jan 2020
5. Nissan: Nissan IDS concept: Nissan's vision for the future of EVs and autonomous driving. Available

(online): <https://usa.nissannews.com/en-US/releases/nissan-ids-concept-nissan-s-vision-for-the-future-of-evs-and-autonomous-driving> (2015). Last Accessed on: 6 Jan 2020

6. Daimler: The Mercedes-Benz F 015. Available (Online): [mercedes-benz.com/en/innovation/research-vehicle-f-015-luxury-in-motion/](https://www.mercedes-benz.com/en/innovation/research-vehicle-f-015-luxury-in-motion/) (2015). Last Accessed on: 6 Jan 2020
7. Rasouli, A., Tsotsos, J.K.: Autonomous vehicles that interact with pedestrians: a survey of theory and practice. In: IEEE transactions on intelligent transportation systems (2019). (<https://ieeexplore.ieee.org/abstract/document/8667866>) <https://doi.org/10.1109/TITS.2019.2901817>
8. Ziegler, J., et al.: Making Bertha drive – an autonomous journey on a historic route. IEEE Intell. Transp. Syst. Mag. 6(2), 8–20 (2014)
9. Böhm, G.: Future of automotive headlamps – light for sensors. In: Proceedings of the 13th International Symposium on Automotive Lightning, Herbert Utz Verlag GmbH, München (2019)
10. Schüler, S., et al.: Integration of a melanopic-light-unit into a passenger car – initial results from a field study. In: Proceedings of the 13th International Symposium on Automotive Lightning, Herbert Utz Verlag GmbH, München (2019)
11. Hamm, M.: Matrix is everywhere: front, rear and interior lighting goes digital. In: Proceedings of the 12th International Symposium on Automotive Lightning, Herbert Utz Verlag GmbH, München (2017)
12. Anjali, D.V., Lokesh, J., Mulla, S.: Human centric lighting: light for health and wellbeing, pp. 10–13 (2017) (<http://eprints.manipal.edu/149680/>) Details: Conference or Workshop Item (Paper)

Autonomous Vehicles – Automated Vehicles

► Autonomous Vehicle (AV) Lighting Systems

Autosomal Dominant Inherited Color Vision Deficiency

► Tritanopia

AV

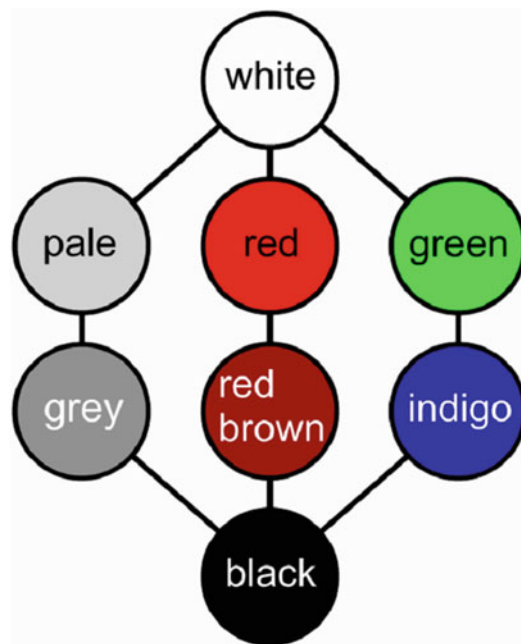
► Autonomous Vehicle (AV) Lighting Systems

Avicenna, Abū ‘Alī al-Ḥusayn ibn ‘Abd Allāh ibn Sīnā

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran



Iran’s Abu ‘Alī al-Husayn ibn ‘Abd Allāh ibn Sīnā (Latinized name: Avicenna) was a Persian polymath and probably the most influential natural philosopher in all of Islamic history [1]. He was born around 980 AD near Bukhara (present-day Uzbekistan) and died in Hamadan (Iran). Already at the age of 16, famous physicians worked under his direction, and at the age of 18, Ibn Sina mastered the contemporary knowledge of the various sciences [2]. His massive *Qanun fi-‘l-tibb* (Canon of Medicine) would be the main medical text in the East and West until the seventeenth century.

The *Kitab al-Shifa* (The cure [of ignorance]) is an equally immense four-part encyclopedia on mathematics, physics, and metaphysics [3, 4]. For several centuries this would be the main text on Aristotelian natural philosophy. Here, Ibn Sina

displayed his Aristotelian sympathies mainly by refuting alternative theories. For theories on light and color, this meant that Ibn Sina rejected the so-called extramission theories advocated by Ptolemy and Euclid. Ibn Sina gave good arguments why it is absurd to assume that vision occurs by visual rays emerging from the eye. Instead, Ibn Sina explained vision in terms of forms transmitted from the visible object to the eye (i.e., by intromission). In this respect, he followed Aristotle just like Ibn Rushd [5]. It would be Ibn al-Haytham, a contemporary of Ibn Sina, who would formulate a successful fusion of the optical theories of Ptolemy, Euclid, Galen, and Aristotle. It would be an intromission theory.

The *Kitab al-Shifa* contains a full chapter on color [6, 7]. In chapter 4 section 3, Ibn Sina was the first to break with the Aristotelian notion that all colors can be ordered along a one-dimensional line, writing:

... the fact that white gradually passes to black by three paths. The first is via pale [...], at first it progresses to pale, from there to grey, and continuing in this manner until black is obtained [...]. There is also another path proceeding [from whiteness] toward red, and from there to red brown, thereafter to black. The third path is the one going to green, from there to indigo, thereafter to blackness. [4, 7]

Thus, Ibn Sina introduced what in modern terms would be called a two-dimensional color order (cf. the illustration above) [8]. It would become known in Christian Europe as well, for example, through its inclusion in the *Speculum maius* from Vincent de Beauvais (1244) [7]. Ibn Sina's texts on color were also widely discussed in the Islamic world. When more than two centuries after Ibn Sina's death a student of the astronomer al-Tusi formulated questions about Ibn Sina's color ordering, al-Tusi further elaborated on this theory [9].

In his commentary, Ibn Sina also criticized the Aristotelian ideas on color mixing. For example, he disagreed with the Aristotelian tradition that claimed that green is composed of red and purple. According to Ibn Sina, a composition of red and purple does not produce green because from mixing red with purple, a color is produced that is brighter than purple but more purple than bright

red. Instead, according to Ibn Sina green is formed by a mixture of yellow, black, and indigo blue [10]. Further, Ibn Sina made a clear distinction between the brightness of a light source (*lux*, in the Latin translation of Ibn Sina's work) and the "splendor" shining from an object (*lumen*) [11].

Finally, Ibn Sina was highly critical about what was known about the cause of the rainbow. He wrote that he was "not satisfied with what our friends the Peripatetics [i.e., Aristotle and his followers] have to say about the rainbow." He continued by reporting some of his observations that cannot be explained by that theory, only to conclude that he himself also had nothing to add, except by a suggestion that the cause of these colors might be inside the eye of the observer [5].

References

1. Lindberg, D.C.: Theories of Vision from al-Kindi to Kepler. University of Chicago Press, Chicago (1976)
2. Anawati, G.C.: Ibn Sina. In: Gillespie, C.C. (ed.) Dictionary of Scientific Bibliography, pp. 494–498. Scribner's Sons, New York (1970)
3. Ibn Sina: Kitab al-Shifa. In: Muntasir, A., Zayid, S., Ismail, A., Madkur, I. (eds.) at-Tabi'iyat 5; Al-ma'adin wa-l-athar al-'ulwiyya. Cairo (1964). Also Qasim, M., Madkur, I. (eds.) at-Tabi'iyat 4; Fi l-'al wa-l-infi'alat. Cairo (1969). Latin translation: van Riet, S. (ed.) Avicenna Latinus: Liber de Anima I-II-III. Édition critique de la traduction latine médiévale. Brill, Louvain-Leiden (1972)
4. Bakos, J.: Psychologie d'Ibn Sina (Avicenne) d'après son oeuvre ash-Shifa. Académie Tchèque des Sciences, Prague (1956). Vol. 1, 105–113; Vol. 2, 75–81
5. Sabra, A.I.: Optics, Islamic. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, pp. 240–247. Scribner's Sons, New York (1982–1989)
6. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) Encyclopedia of Islam. Brill, Leiden (1991)
7. Kuehni, R.G., Schwarz, A.: Color Ordered – A Survey of Color Order Systems from Antiquity to the Present. Oxford University Press, Oxford (2008)
8. Kirchner, E.: Color theory and color order in medieval Islam: a review. Col. Res. **40**, 5–16 (2015)
9. Kirchner, E., Bagheri, M.: Color theory in medieval lapidaries: Nishaburi, Tusi and Kashani. Centaurus. **55**, 1–19 (2013)
10. van Campen, M.: De regenboog bij de Arabieren. University of Utrecht, Utrecht (1988) MSc Thesis, 95
11. Lindberg, D.C.: Optics, Western European. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, vol. 9, pp. 247–253. Scribner's Sons, New York (1989)

B

Balance

► [Color Harmony](#)

Bartleson, C. James

Robert W. G. Hunt
Department of Colour Science, University of
Leeds, Leeds, UK



C. James Bartleson was an American color scientist who made very important contributions to colorimetry and visual science.

Bartleson graduated as an Associate of Photographic Science at the School of Photography at Rochester Institute of Technology (RIT), Rochester NY, USA, in 1951. He received a Ph.D. at the City University, London, England, in 1977.

After high school, he enlisted in the marines and served as a frogman using photography. After his discharge in 1948, he spent a year doing freelance aerial photography and then enrolled in the School of Photography at RIT to obtain his degree of Associate of Photographic Science in 1951. There followed a 3-year period working on the new Ansco Plenachrome system at the Pavelle Color Corporation in New York City.

In 1952, he moved to Rochester, NY, to start a career with the Eastman Kodak Company. His first assignment was in the Color Control Department under Ralph Evans but his flair for research was noticed and in 1957 he moved to the Physics Division where he started his research on color and tone reproduction which led to the publication of a series of very significant papers, one of the most outstanding, in 1967, coauthored with Ed Breneman, being on the effect of light and dark surrounds on apparent contrast. His papers attracted various awards, and he was one of the first recipients of the C.E.K. Mees Award. He was also author, coauthor, or editor of several books, including the five volume series on "Optical Radiation Measurements." He was also much sought after as a lecturer, and his award of the British Color Group's Newton Medal was accompanied by a masterly presentation.

In 1967, his outstanding reputation led him to be chosen to establish a research facility for the MacBeth Company, later the MacBeth Color and Photometry Division of the Kollmorgen Corporation, in Newburgh, NY.

When his time at Kollmorgen came to an end in 1974, the following year, at the then age of 45, he enrolled in the School of Ophthalmic Optics and Visual Science at The City University, London, England, and obtained his Ph.D. in 1977. With typical enthusiasm, he constructed an ingenious apparatus for studying the effects of adaptation on perception, and the resulting thesis was of outstanding quality.

He then returned to Rochester, NY, to resume his career at Kodak, where he soon established a reputation as a most valuable consultant for other members of the staff.

He was the President of the AIC (Association Internationale de la Couleur) from 1978 to 1981, reviving it from a moribund state, and he took part in several committees of the CIE (Commission Internationale de l'Éclairage).

His main interests included optics, vision, photometry, colorimetry, color perception, color applications, and color photography. His leisure interests also included photography and travel.

Further Reading

1. Burnham, R.W., Haynes, R.M., Bartleson, C.J.: *Color: A Guide to Basic Facts and Concepts*. Wiley, New York (1963)
2. Grum, F. (ed.): *Optical Radiation Measurements*. Academic, New York (1984)
3. Grum, F., Becherer, R. J.: *Optical Radiation Measurements. volume 1 Radiometry*. Academic, New York (1979)
4. Grum, F., Becherer, R. J.: *Optical Radiation Measurements. volume 2 Color Measurement*. Academic, New York (1980)
5. Mielenz, K. D. (ed.): *Optical Radiation Measurements. volume 3 Measurement of Photoluminescence*. Academic, New York (1982)
6. Budde, W.: *Optical Radiation Measurements. volume 4 Physical Detectors of Optical Radiation*. Academic, New York (1983)
7. Grum, F., Becherer, R. J.: *Optical Radiation Measurements. volume 5 Visual Measurements*. Academic, New York (1984)

Basic Color Categories

► [Color Categorization: Patterns and Mechanisms of Evolution](#)

Basic Color Terms

► [Color Categorization: Patterns and Mechanisms of Evolution](#)

Basic Colors

► [Primary Colors](#)

Batik

Traci A. M. Lamar
Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Synonyms

[Resist printing; Wax-resist dyeing](#)

Definition

Batik is a process for creating patterned textiles by applying wax in defined areas to form a dye resist, then subsequently dyeing the fabric. A textile created using this process is also referred to as a batik.

Background, Process, and Aesthetic Properties

There is no general agreement in the textile coloration field regarding where the practice of batik originated. Early examples of batik have been

found in India, Central Asia, the Far East, and the Middle East [1]. The techniques and tools used to accomplish batik varied from region to region and still do. In many parts of the world batik is an important component of the cultural heritage and this is reflected in the often symbolic colors, motifs, and designs used in the traditional batiks. Many sources identify the term batik as an Indonesian word deriving from the Javanese *amba*, meaning “to write” and *titik*, meaning “dots.” Creating hand-drawn, detailed batiks is an exacting process and a lengthy one. Indonesian batik has long been, and continues to be, widely regarded as reflecting the most advanced practice and the highest standards, especially on the island of Java. In 2009 Indonesian batik was inscribed to UNESCO’s Representative List of the Intangible Cultural Heritage of Humanity [2]. Therefore, the methods, tools, and examples for Indonesian batik are used to illustrate the batik process.

Batik is typically characterized as a form of resist printing, or resist dyeing, due to the pattern being formed during the coloration process by an applied dye resist. In the batik process, hot liquid wax is applied to a textile to create areas of resist that prevent dye penetration and areas that retain their original color during subsequent dyeing.

The textiles used for Indonesian batiks are typically finely woven cotton. To begin, the fabric is stretched across a frame made of wood or bamboo called a “gawangan” [3] to provide support. For large batiks, the fabric may be stretched across a padded table instead. Paraffin wax, beeswax, or a mixture of these is commonly used in the batik process. Resin may be added to the wax. The specific desired effect defines the best wax blend to use. Indonesian batik traditionally uses beeswax, and paraffin wax is mixed in to enhance the cracking effect. When fresh, the wax is translucent and yellow but may look dark or brown after being reused many times [4]. The tjanting or canting is a traditional tool used to apply wax in the batik process (Fig. 1). The tjanting consists of a copper cup to hold the hot wax attached to a bamboo handle.

The copper cup has one or more spouts through which hot wax can flow. The tjap or cap, a block comprised of strips of cut copper shaped into



Batik, Fig. 1 A batik tjanting for applying wax by hand drawing. (Image credit: Gregg Museum of Art & Design, Raleigh NC; Gift of Helen S. Underhill, 2001.035.031)

detailed patterns and soldered with a handle on the back, can be used to stamp hot wax onto the textile. The tjap allows wax to be applied much more quickly, and also for the easy creation of repeating patterns. The tjap can also be used with the tjanting for quicker wax application, and is especially useful for large batiks. Some batik artists also use a tool made by wrapping fabric around a stick to brush wax onto large areas of the textile. In the highest quality batiks, wax is applied to both sides of the fabric guaranteeing batiks with clean, sharp designs. Waxing both sides of the fabric lengthens the process of creating a batik, so it is not always done. Maintaining a constant wax temperature is also important to the quality of the batik because wax temperature impacts how much the wax runs when applied.

After the application of the wax resist (Fig. 2), the textile is dyed. The resist created by the wax may crack during the coloration step, allowing dye to penetrate the waxed areas and creating a crackled effect in those areas. Prior to dyeing, cracking can be increased if desired by folding, twisting, or otherwise manipulating the cloth after the wax has cooled and hardened. Traditional Indonesian batik uses natural dyes, although in areas of Indonesia use of synthetic dye has shortened the dyeing process and contributed to richer and more varied colors than the traditional indigo blues, browns and reds. The initial dye bath contains the lightest color needed in the final design, and waxed areas of the fabric resist dye penetration remaining white. The dyed fabric is then washed. After cleaning, waxing is repeated but



Batik, Fig. 2 Full view (*top*) and detail view (*bottom*) of batik fabric before dyeing, with wax designs still on the fabric. The design was created by both hand drawn and stamped wax application and includes traditional

Indonesian wayang puppet figures. (Image credit: Gregg Museum of Art & Design, Raleigh NC; Gift of Dorothy H. Pope, 1997.008.002)



Batik, Fig. 3 Full view (*top*) and detail view (*bottom*) of batik with design featuring puppet-like figures similar to undyed waxed fabric in Fig. 2. (Image credit: Gregg

Museum of Art & Design, Raleigh NC; Gift of Dorothy H. Pope, 1997.018.001)

this time areas to remain white and areas to remain the color of the initial dyeing are waxed. The fabric is then placed in a second dye bath containing the next lightest color in the design.

This process of waxing, dyeing, and washing is repeated until all colors of the design have been achieved and a multicolored image is revealed on the fabric. When the dyeing process is complete,

Batik, Table 1 Aesthetic properties of batik

Cracking or veining	Dye penetration of small cracks in the wax resist during dyeing creates veined patterns in the batik when the wax is removed
Flowing, liquid quality	Lines and details with a flowing, liquid quality are found in batiks, due to the hot, liquid wax being applied to create the design
Halos	In many designs, slightly darker or lighter halos are found around pattern motifs and lines
Saturated color	Because batik involves dyeing the textile, areas that are allowed to absorb dye are saturated through with color
Design on both face and back	Because wax is often applied to both sides of the textile, the batik motifs, lines, and patterns are created on both sides



Batik, Fig. 4 Batik from Java, Indonesia with tjap stamped motifs and traditional dotted patterning. (Image credit: Gregg Museum of Art & Design, Raleigh NC; Anonymous gift, 1986.005.006)

all of the wax resist must be removed from the base fabric. This is accomplished through boiling the batik in water containing a detergent. Finally, the batik is rinsed and dried (Fig. 3).

Contemporary batik differs from traditional, more formal types. Batik artists use a variety of textile materials and may combine batik with other surface design methods such as etching or discharge printing. Contemporary batiks are created with traditional tools, nontraditional tools, and contemporary versions of traditional tools. For example, an electric tjanting allows contemporary artists to adjust the flow of the wax and the wax temperature to be maintained at a constant level, facilitating high-quality wax effects. Lawrence [5] identified five aesthetic properties characterizing contemporary batik, which are described in Table 1. Some of these properties can be seen in the Fig. 4 detail view. It is important to note that prevalence of and preference for veining in batik is culturally based, and cracking is traditionally undesirable in Javanese batik but quite valued in some other cultures.

Cross-References

► [Resist Printing](#)

References

1. The Batik Guild. The History of Batik. <https://www.batikguild.org.uk/batik/history-of-batik> (1999). Accessed 8 Aug 2019
2. Indonesian Batik. United Nations Educational, Scientific and Cultural Organization. <https://ich.unesco.org/en/RL/indonesian-batik-00170> (2009). Accessed 8 Aug 2019
3. Spee, M.: Traditional & Modern Batik. Kangaroo Press, Hong Kong (1982)
4. Asian Art Museum Chong-Moon Lee Center for Asian Art and Culture. Batik of Java: A Centuries Old Tradition. <https://education.asianart.org/explore-resources/video/batik-java-centuries-old-tradition-0> (2017). Accessed 30 Aug 2019
5. Lawrence, G. Digital Printing and Traditional Surface Design Techniques. Unpublished M.S. Thesis, NC State University. <https://repository.lib.ncsu.edu/handle/1840.16/2948> Accessed 9 Aug 2019

Bayesian Approaches to Color Category Learning

Thomas L. Griffiths¹ and Noga Zaslavsky²

¹Departments of Psychology and Computer Science, Princeton University, Princeton, NJ, USA

²Department of Brain and Cognitive Sciences and Center for Brains Minds and Machines, MIT, Cambridge, MA, USA

Synonyms

[Ideal observer models of color category learning](#);
[Rational models of color category learning](#)

Definition

Bayesian approaches to color category learning formalize learning as a problem of Bayesian inference, requiring the learner to form generalizations that go beyond observed examples of members of a category. This formal framework can be used to make predictions about both individual judgments and how populations form color categories.

Color Category Learning

One of the challenges that children face as they acquire a language is discovering how words are used to refer to different colors. While human languages demonstrate variation in how they partition the space of colors, there are also clear regularities in the kinds of systems of color categories that are used [1, 2]. This raises two important questions: How might color categories be learned? And how might regularities in systems of color categories across languages be explained?

Learning color categories is an inductive problem, requiring learners to make an inference from labeled examples of colors to a full system of color categories. As in other domains of perception [3], an “ideal observer” model can be used to

explore the optimal solution to this problem. Let h denote a hypothesis about a possible system of color categories and d the observed data – a set of labeled examples (such as “This color is blue, and this color is yellow”). If learners represent the degree of belief in the truth of each hypothesis with a probability, $P(h)$, then the ideal solution to the problem of updating these beliefs in light of the data d is provided by Bayes’ rule:

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h'} P(d|h')P(h')}, \quad (1)$$

where $P(h|d)$ (known as the posterior probability, in contrast to the prior probability $P(h)$) indicates the degree of belief assigned to h after observing d and $P(d|h)$ (known as the likelihood) indicates the probability of observing d if hypothesis h were true.

The sum in the denominator of Bayes’ rule ranges over all possible hypotheses and ensures that $P(h|d)$ is a valid probability distribution, summing to 1. The key idea behind Bayes’ rule can be obtained by ignoring this constant and simply inspecting the numerator: The new beliefs of the learner result from combining the previous beliefs, captured in the prior distribution $P(h)$, with the probability of the observed data under each hypothesis, expressed by the likelihood $P(d|h)$. The prior distribution captures the expectations of the learner, but also indicates which hypotheses are easy or hard to learn. A hypothesis that has low prior probability requires stronger evidence (in the form of a higher likelihood) to end up with a high posterior probability and so will be harder to learn. The prior distribution thus provides a way of encoding the perceptual or learning biases of the learner, favoring some hypotheses over others.

The Bayesian approach to modeling learning has proven successful in accounting for human behavior in a wide range of tasks [4]. In particular, Bayesian models have been used to account for how people learn new concepts and new words. Tenenbaum and Griffiths [5] presented an account of how people form generalizations from examples, such as inferring what other numbers might

belong to a set when told that the set contains 2, 8, and 64. Under this account, hypotheses correspond to possible sets of numbers, and the likelihood is obtained by calculating how likely it is that the examples would be observed if they were sampled at random from this set. Xu and Tenenbaum [6] showed that a closely related model captured how children learned nouns corresponding to sets of objects, such as determining the appropriate referent of words corresponding to “Dalmatian” or “dog.” These results suggest that a Bayesian approach might also be fruitful for explaining the acquisition of terms for color categories.

Dowman [7] presented a model that took exactly this approach, providing a Bayesian account of color category learning. In this model, the space of colors is reduced to a one-dimensional ring of hues. Each color category then corresponds to an interval on this ring, picking out a set of adjacent colors. Labeled examples of color categories are assumed to be sampled from the categories at random, with a small probability of an error taking place. This makes it possible to calculate the probability of any observed set of labeled examples for each candidate interval from which they might be drawn, providing the likelihood $P(d|h)$. Bayes’ rule can then be used to compute a posterior distribution over possible intervals for each color category. The probability that a color that has not previously been labeled belongs to that category is then obtained by summing the probability of all intervals that contain that color under the posterior distribution. Dowman demonstrated that this model made reasonable inferences for simplified versions of the systems of color categories from real languages, such as Urdu.

The predictions that Dowman’s model makes about learning of color categories have not been directly tested with human learners, but results in other domains and with other species provide support for this approach. As noted above, Xu and Tenenbaum [6] found that a very similar model accounted well for the generalizations that children made in learning novel words describing sets of objects. In addition, Jones, Osorio, and Baddeley [8] found that poultry chicks form

generalizations about colors in a conditioning task that is consistent with a model based on that of Tenenbaum and Griffiths [5].

Cultural Transmission of Color Category Systems

The results summarized so far indicate how Bayesian models might be used to explain learning of color categories. The same models also have the potential to provide insight into why regularities exist in the systems of color categories that appear across human languages. Dowman [7] explicitly had this goal in mind in defining his Bayesian learning model, which was used as a component in a simulation of the cultural transmission of systems of color categories. Dowman's aim was to investigate the consequences of cultural transmission of systems of color categories among a set of agents that used a realistic approximation to human learning. He found that cultural transmission by Bayesian agents produced systems of color categories with properties similar to those seen across human languages, providing a potential explanation for the source of those regularities.

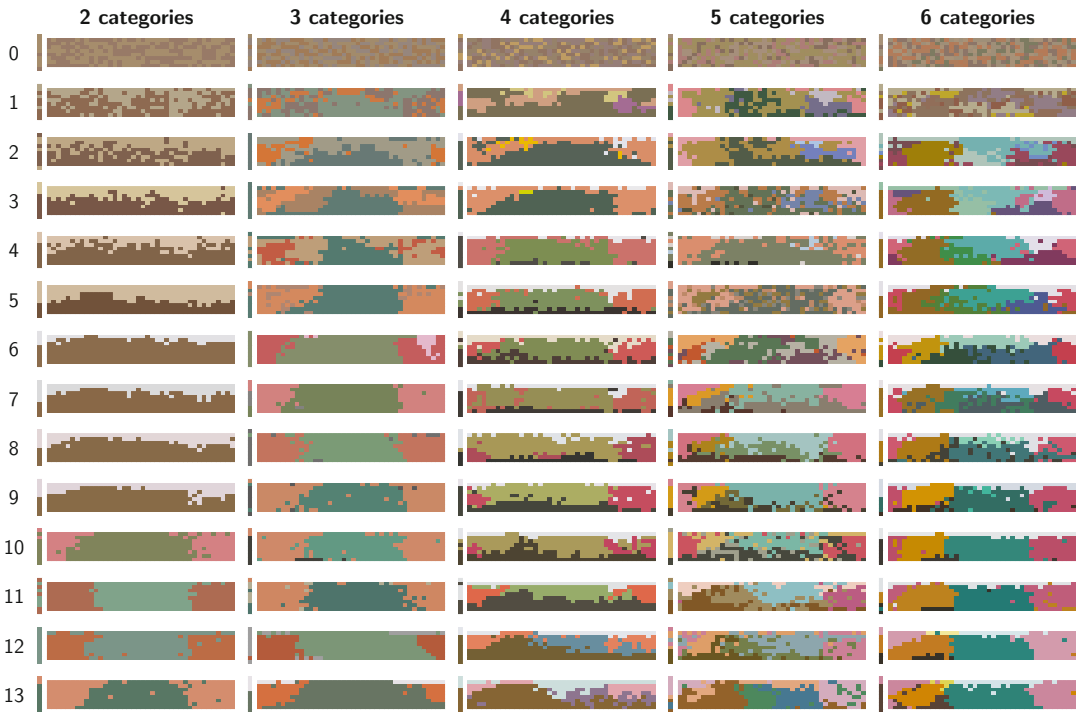
Dowman's simulation of cultural transmission was an instance of a more general approach to exploring the origins of different kinds of structure in human languages, known as iterated learning [9]. In an iterated learning model, a set of agents each learn from some observed data and generate data that is observed by other agents. The simplest case is where the agents form a chain, with each agent learning from data generated by the previous agent and generating data that is provided to the next agent. Griffiths and Kalish [10] showed that when this form of iterated learning is carried out by Bayesian agents who all have the same prior, the hypotheses considered by those agents eventually converge to a distribution that matches the prior distribution. More precisely, the probability that an agent selects a hypothesis h converges to the prior probability of that hypothesis $P(h)$ as the chain gets longer.

The results of Dowman [7] and Griffiths and Kalish [10] raise an interesting question: Can the

regularities seen in systems of color categories across human languages be accounted for by cultural transmission producing convergence on a shared prior distribution? To explore this question, Xu, Dowman, and Griffiths [11] conducted an experiment in which human participants simulated cultural transmission by iterated learning. Each participant was given some examples of colors from novel categories and then asked to generalize to a larger set of colors. The generalization responses were then recorded and used to generate the examples that were seen by the next participant. Figure 1 shows examples of color category systems produced by chains of participants in this experiment. Each chain was initialized with data generated from a random partition of color space into labeled categories. The number of allowed categories varied across chains, ranging from two to six categories. Xu et al. [11] found that over time, the systems of color categories converged to forms that were consistent with the regularities seen across human languages [2].

These results support the idea that cultural transmission and perceptual or learning biases of the kind that might be captured by a prior distribution in a Bayesian model may be sufficient to explain the origins of cross-linguistic regularities in systems of color categories. However, much remains unknown about the precise characterization of the human prior distribution over systems of color categories and whether the structure of that distribution can capture cross-linguistic variation. One approach to this question was suggested by Carstensen, Xu, Smith, and Regier [12], building on the idea that languages are shaped by pressure for informative communication. Carstensen et al. [12] showed that the iterated learning experiment of Xu et al. [11] produced systems of color categories that become highly informative, in addition to becoming more similar to actual color category systems of human languages [11]. They concluded that cultural transmission with a bias toward informative communication may explain the constrained cross-linguistic variation seen in color category systems.

Carr, Smith, Culbertson, and Kirby [13] proposed an alternative explanation for these results.



Bayesian Approaches to Color Category Learning, Fig. 1 Examples of color category systems produced by chains of human learners in the iterated learning experiment of Xu et al. [11] (figure adapted from [11]). Each column corresponds to one cultural transmission chain with 13 participants (rows) and 2–6 categories. All systems are plotted against the World Color Survey stimulus palette [2], which contains 10 achromatic color chips and 320 maximally saturated color chips arranged by hue

(horizontal axis) and lightness (vertical axis). Each category is color-coded by its color centroid in CIELAB space and covers all the chips in the palette that were assigned to it. The chains were initialized with a random system, shown at iteration 0. Over time, the systems produced by the learners become more similar to actual systems across languages [11], capturing observed cross-linguistic regularities.

They argued that a human prior distribution over category systems is better characterized by a bias toward simplicity for learning rather than informativeness. Carr et al. [13] suggested formal definitions for a simplicity prior and an informativeness prior and then used these prior distributions for model simulations of iterated learning with Bayesian agents. They demonstrated that the simplicity model can produce structured category systems that are also highly informative, which may give an impression of a bias toward informativeness even in cases where it is not actually present. In addition, Carr et al. [13] tested these models on human iterated learning data and found stronger empirical support for a bias toward simplicity. Their results, however,

are based only on synthetic domains that are qualitatively different than color. Therefore, it remains unclear to what extent simplicity, informativeness, or possibly both, may characterize human biases involved in learning systems of color categories.

Bayesian approaches to color category learning can be used to explore questions about how children might learn how their language partitions the space of colors, why regularities are seen in systems of color categories across languages, and how color categories may evolve over time. However, many important questions remain open for future research. One fundamental question is how well Bayesian models can capture the generalizations that real human children make when

learning color categories. Another fundamental question for future research is how communication between two or more agents within a generation may influence the patterns that emerge in culturally transmitted systems of color categories. Communication is believed to play an important role in the cultural evolution of language [14], in addition to learning. Furthermore, it has been argued that systems of color categories evolve along an information-theoretically optimal trajectory of communication systems involving Bayesian agents [15]. This information-theoretic approach captures much of the regularities and variation seen in systems of color categories across languages, suggesting a potential theoretical link between communication and learning in Bayesian models of color categories.

Cross-References

- [Berlin and Kay Theory](#)
- [Color Categorical Perception](#)
- [World Color Survey](#)

References

1. Kay, P., Maffi, L.: Color appearance and the emergence and evolution of basic color lexicons. *Am. Anthropol.* **101**(4), 743–760 (1999)
2. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. Center for the Study of Language and Information, Stanford (2009)
3. Kersten, D., Mamassian, P., Yuille, A.: Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**(1), 271–304 (2004). <https://doi.org/10.1146/annurev.psych.55.090902.142005>
4. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D.: How to grow a mind: Statistics, structure, and abstraction. *Science*. **331**(6022), 1279–1285 (2011). <https://doi.org/10.1126/science.1192788>
5. Tenenbaum, J.B., Griffiths, T.L.: Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* **24**(4), 629–640 (2001). <https://doi.org/10.1017/S0140525X01000061>
6. Xu, F., Tenenbaum, J.B.: Word learning as Bayesian inference. *Psychol. Rev.* **114**(2), 245–272 (2007). <https://doi.org/10.1037/0033-295X.114.2.245>
7. Dowman, M.: Explaining color term typology with an evolutionary model. *Cogn. Sci.* **31**(1), 99–132 (2007). <https://doi.org/10.1080/03640210709336986>
8. Jones, C.D., Osorio, D., Baddeley, R.J.: Colour categorization by domestic chicks. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **268**(1481), 2077–2084 (2001). <https://doi.org/10.1098/rspb.2001.1734>
9. Kirby, S.: Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* **5**(2), 102–110 (2001)
10. Griffiths, T.L., Kalish, M.L.: Language evolution by iterated learning with Bayesian agents. *Cogn. Sci.* **31**(3), 441–480 (2007). <https://doi.org/10.1080/15326900701326576>
11. Xu, J., Dowman, M., Griffiths, T.L.: Cultural transmission results in convergence towards colour term universals. *Proc. Roy. Soc. B: Biol. Sci.* **280**(1758), 20123073 (2013). <https://doi.org/10.1098/rspb.2012.3073>
12. Carstensen, A., Xu, J., Smith, C., Regier, T.: Language evolution in the lab tends toward informative communication. In: Noelle, D.C., Dale, R., Warlaumont, A.S., Yoshimi, J., Matlock, T., Jennings, C.D., Maglio, P.P. (eds.) *Proceedings of the 37th annual meeting of the Cognitive Science Society*. Cognitive Science Society, Austin (2015)
13. Carr, J.W., Smith, K., Culbertson, J., Kirby, S.: Simplicity and informativeness in semantic category systems. *Cognition*. **202**, 104289 (2020). <https://doi.org/10.1016/j.cognition.2020.104289>
14. Tamariz, M.: Experimental studies on the cultural evolution of language. *Ann. Rev. Linguist.* **3**(1), 389–407 (2017). <https://doi.org/10.1146/annurev-linguistics-011516-033807>
15. Zaslavsky, N., Kemp, C., Regier, T., Tishby, N.: Efficient compression in color naming and its evolution. *Proc. Natl. Acad. Sci.* **115**(31), 7937–7942 (2018). <https://doi.org/10.1073/pnas.1800521115>

Bend Light – Cornering Light

- [Automotive Front Lighting System \(Status 2020\)](#)
-

Benham's Disc

- [Fechner's Colors and Benham's Top](#)

Berlin and Kay Theory

C. L. Hardin

Department of Philosophy, Syracuse University,
Syracuse, NY, USA

a variety of non-hue senses, taking on predominantly hue sense during the Middle English period. *Purple* and *orange* were borrowed from Latin and Arabic with a hue sense from the beginning, and *pink* emerged sometime in the sixteenth century.

Definition

The Berlin-Kay theory of basic color terms maintains that the world's languages share all or part of a common stock of color concepts and that terms for these concepts evolve in a constrained order.

Basic Color Terms

In 1969 Brent Berlin and Paul Kay advanced a theory of cross-cultural color concepts centered on the notion of a *basic color term* [1]. A basic color term (BCT) is a color word that is applicable to a wide class of objects (unlike *blonde*), is monolexemic (unlike *light blue*), and is reliably used by most native speakers (unlike *chartreuse*). The languages of modern industrial societies have thousands of color words, but only a very slender stock of basic color terms. English has 11: *red, yellow, green, blue, black, white, gray, orange, brown, pink, and purple*. Slavic languages have 12, with separate basic terms for light blue and dark blue.

In unwritten and tribal languages the number of BCTs can be substantially smaller, perhaps as few as two or three, with denotations that span much larger regions of color space than the BCT denotations of major modern languages. Furthermore, reconstructions of the earlier vocabularies of modern languages show that they gain BCTs over time. These typically begin as terms referring to a narrow range of objects and properties, many of them noncolor properties, such as succulence, and gradually take on a more general and abstract meaning, with a pure color sense.

The development of English BCTs is typical [2]. The terms *red, black, white, gray, red, green, yellow, and brown* have roots that go back to Old English and before. All of them except for *red* had

The Original Berlin-Kay Theory

These considerations invite two questions. Are there similarities in the ways that different languages with the same number of BCTs carve up color space? Is the order in which languages acquire BCTs constrained? In their influential 1969 *Basic Color Terms* [1], Berlin and Kay answered both questions in the affirmative and proposed a sequence in which the number of BCTs in a language predicts what those terms would be. According to the scheme, a Stage I language will contain terms for black and white. In Stage II, these are joined by a word for red. In Stages III and IV, either a word for green appears, followed by a word for yellow, or else a word for yellow appears first, followed by a word for green. A word for blue arises in Stage V, followed by a word for brown in Stage VI. Thereafter, in Stage VII, words for pink, purple, orange, and gray appear in no particular order.

The World Color Survey and the Revised Berlin-Kay Theory

The 1969 sequence has been enshrined in many textbooks, but it has since been significantly modified and conceptually refined. The authors and their collaborators improved their methodology and greatly extended the scope of their samples with the 1976 *World Color Survey* (WCS) [3] published with analysis and interpretation in 2009, along with work by others, notably MacLaury's 1997 *Color and Cognition in Mesoamerica* [4]. For the WCS, Protestant missionaries from the Summer Institute of Linguistics collected data from 25 monolingual speakers of each of 110 unwritten minor and tribal languages from 45 different language families. Respondents

were asked to name 320 Munsell color chips of maximum chroma and 8 value levels in addition to 10 achromatic chips, presented in random order. After two such naming sessions, they were presented with a miniature version of all 330 samples arrayed in standard order and asked to pick out focal (best) examples of each of the named colors. The resulting dataset [5] permitted much more sophisticated statistical analysis, which broadly – but not completely – confirmed Berlin and Kay’s original conclusions. However, in the years between the appearance of *Basic Color Terms* and the publication of the World Color Survey, the original sequence was replaced by a substantially more complex developmental scheme, as well as a reconceptualization of what that development consisted in.

The original sequence was understood as a succession of encodings of basic colors in which a language would acquire a term for red, and a bit later yellow or green, then blue and so on. It soon became clear that this wouldn’t do, since in languages with fewer BCTs, the reference for each color term would encompass a correspondingly larger region of color space – a megacategory – so the addition of each new BCT should be understood as breaking up a megacategory into smaller categories. For instance, in a Stage I language with just two color categories, referring to one category as “white” and to the other as “black” is misleading, since the native category names extend across a broad range of hue samples. The more accurate gloss would be “warm/light” for the “white” category and “dark/cool” for the “black” category. In WCS notation, Stage I is represented as [W/R/Y, Bk/G/Bu]. In Stage II, the first megacategory breaks up, yielding [W, R/Y, Bk/G/Bu], and so on, culminating in Stage V, consisting of [W, R, Y, G, Bu, Bk]. The remaining BCTs appear in no fixed order.

A well-established basic megacategory in many North American and Mesoamerican Indian as well as some African languages is a “grue” term that covers the region that English speakers would separate into a blue and a green category. English, in turn, has its own megacategory. Modern Russian speakers break up our blue category with two BCTs: *goluboi*, light blue, and *sinji*, dark blue.

Constraints on Color-Category Formation

A fundamental question now arises. If the development of basic color categories is constrained in various ways, what are the bases of these constraints? One striking feature of the sequence is that as megacategories break up, categories whose foci lie close to the six elementary Hering colors get named before categories focused near the binary colors such as orange. Furthermore, the focal choices for each of these categories tend to cluster around average unique hue choices. Studies have found the variability of focal choices among the speakers of a given language to be substantially greater than the variability of average focal choices across languages with a comparable number of BCTs. This suggests that the elementary colors are perceptually salient for all language users, making them more likely to be named [6].

Alternative Explanations

An alternative view, first advanced by Jameson and D’Andrade [7], starts with the observation that the outer surface of the Munsell solid that was used in the collection of the WCS data is lumpy. Regions of high chroma are “mountains” of greater perceptual prominence than “valleys” of lower chroma in which the samples are less readily discriminable from their neighbors. This salience makes them more readily nameable.

Kay and Regier [8] constructed a CIELAB representation of the WCS Munsell samples and used them to define a dispersion metric. This enabled them to determine individual and group centroids for every category of each WCS language. They found a strong tendency of category centroids to cluster across languages. Subsequently, they used the same CIELAB representation to formalize and test the ideas of Jameson and D’Andrade [9]. Defining similarity and dissimilarity measures, they derived a measure of the “well formedness” of a language. By this they mean the extent to which the language maximizes the perceptual similarity of colors within a category and minimizes it across categories. Rotating

the data for each of the WCS languages around the lightness axis showed that in most cases the well formedness of the language was maximized at or very near 0° of rotation. They concluded that the irregularities in perceptual color space along with general principles of category formation could account for the organization of the WCS languages. Well formedness is not overly sensitive to the placement of category boundaries, according to this account, and permits adjustment by linguistic convention.

Other explanations for the Berlin-Kay structure take different approaches. Vantage Theory, offered by Robert MacLaury, is psychological, giving a role to individual perceptual-cognitive differences. By contrast, a theory advanced by Sergej Yendrikhovskij [10] maintains that the structure of color-category systems originates in the statistics of the natural environment. He developed a clustering model that predicted the location, number and order of color categories and tested the model with a database of 630 natural images. The clustering was consistent with the 11 basic color categories of Berlin and Kay. These theories of universal constraints are not necessarily incompatible, since they focus on different aspects of color categorization.

Further Questions

Many questions concerning the evolution and structure of color categories remain to be answered. Why do red terms seem to be the first hue words to appear in color lexicons? Why do grue categories persist for so long in color-category evolution? How are we to understand the appearance of categories in languages from lands as far apart as New Guinea and North America that lump yellow with green instead of with red?

It is very likely that social factors such as trade and conquest drive the evolution of color vocabularies. It seems clear that culture plays a significant role in both the origins and the boundaries of color categories. Exactly what is that role? Can culture override perceptually based constraints? It appears that a proper understanding of even the denotations – let alone the connotations – of a

language's color terms requires a proper grasp of the relative contributions of biological, cultural, and environmental factors [11].

Cross-References

- ▶ [Ancient Color Terminology](#)
- ▶ [Color Categorical Perception](#)
- ▶ [Dynamics of Color Category Formation and Boundaries](#)
- ▶ [Effects of Color Terms on Color Perception and Cognition](#)
- ▶ [Infant Color Categories](#)
- ▶ [Unique Hues](#)
- ▶ [Vantage Theory of Color](#)
- ▶ [World Color Survey](#)

References

1. Berlin, B., Kay, P.: Basic Color Terms. University of California Press, Berkeley (1969)
2. Casson, R.W.: Color shift: evolution of English color terms from brightness to hue. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 224–239. Cambridge University Press, Cambridge (1997)
3. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
4. MacLaury, R.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, Austin (1997)
5. <http://www.icsi.berkeley.edu/wcs/data.html>
6. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R.L., Parmei, G., Dedrick, D. (eds.) *The Anthropology of Color*. John Benjamin, Amsterdam (2007)
7. Jameson, J., D'Andrade, R.G.: It's not really red, green, yellow, blue: an inquiry into perceptual color space. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 224–239. Cambridge University Press, Cambridge (1997)
8. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**(15), 9085–9089 (2003)
9. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. U. S. A.* **104**(4), 1436–1441 (2007)
10. Yendrikhovskij, S.N.: A computational model of color categorization. *Color. Res. Appl.* **26**, 235–238 (2001)
11. Belpaeme, T., Bleys, J.: The impact of statistical distribution of colours on colour category acquisition. *J. Cogn. Sci.* **10**(1), 1–20 (2009)

Berlin-Kay Hypothesis

► [Color Categorization: Patterns and Mechanisms of Evolution](#)

Bezold–Brücke Effect

► [Assimilation](#)

Bilingual Color Naming/ Categories

Prutha S. Deshpande¹ and Nancy Alvarado²

¹Department of Psychology, Ohio State University, Columbus, OH, USA

²Department of Psychology and Sociology, California State Polytechnic University, Pomona, CA, USA

Definition

A bilingual person is a person fluent in more than one language, either from birth or acquired later in life. A multilingual person is fluent in more than two languages. Color naming is the process of assigning color terms to refer to color appearances in the world. Color categories are mental representations of the regions of color space that are designated by particular color terms or names.

Cross-language Comparisons

The influence of culture on color perception and categorization is most often studied by comparing the use of single-word (monolexic) color terms or Basic Color Terms (see entry ► [“Berlin and Kay Theory”](#)) to name color categories across different languages, as in the **World Color Survey** [1, 2]. Color categorization and naming have been studied across a wide range of cultures to explore questions about the universality of basic

color categories and the ways in which color naming may affect color perception. However, such comparisons between groups of individuals speaking different languages in different cultures necessarily confound differences in individual experience and individual perception with the linguistic properties of the languages spoken, and the categorization practices of the cultures themselves, resulting in a difficulty identifying the distinct contribution of each source of influence on color naming behavior. Comparison of the categorization and naming behavior of multilingual or bilingual individuals using each of their distinct languages provides a way of controlling for individual differences so that the influence of culture and linguistics on naming can be more readily determined.

Cross-language Comparisons Within an Individual

Comparisons of multilingual/bilingual individuals performing identical color naming or categorization tasks separately in each of their languages demonstrate the influence of several factors. The sequence in which languages were acquired and the relative proficiency in each language are important and should be measured in studies of multilingual or bilingual color naming. In some cases, a person may know two or more languages because he or she is an immigrant to a new culture, in the process of acquiring a new language, while using the native language or previously acquired languages less frequently. An individual in the process of assimilating to a new culture and language may experience interference from previous languages. He or she may be motivated to forget, or experience pressure to forget, previously acquired languages, affecting naming behavior. Studies comparing bilingual naming to monolingual naming show increased consensus of response among bilinguals, as well as a preference for use of more basic color terms [3, 4]. The same phenomenon has been observed for emotion terms [5]. In Loitšenko’s [6] comparisons of Estonian/Russian bilinguals dominant in either Estonian or Russian, both groups of bilinguals

showed convergence in their responses in tasks of color listing and naming, with consensus however being higher with the language more frequently used. These patterns perhaps reflect a focus on what is shared across languages, or a vocabulary loss affecting the less frequently used, more differentiated terms needed to describe subtleties of color samples. Others have further suggested that immersion in a new culture may affect the availability of environmental referents, or the communicative salience of specific color differences, and thus the availability of previously learned color terms in semantic memory (see entry ► [“Memory Color”](#)) [7]. For these reasons, bilingual color categorization and naming behavior should be compared to both monolingual immigrant and nonimmigrant populations. Few studies have made such comparisons, even in the literature on bilingualism.

In other cases, a person may be a participant in a multicultural society in which different languages are spoken in different contexts and for different purposes, but are generally spoken frequently enough to be refreshed in memory. Code-switching occurs when bilingual or multilingual individuals spontaneously change languages while speaking. The switching suggests that some concept might be better expressed using a particular language or that access to vocabulary is easier in that language. Study of when and how this occurs is an emerging field in linguistics [8]. In comparisons of bilingual color naming, subjects are required to switch languages to perform the same task twice but are constrained to respond in one language. For example, Deshpande [9] examined the color naming behavior and category extensions of Korean-English bilinguals who repeated the same experimental procedures over two sessions, in each of their language modes. While such comparisons can certainly be informative, it is unclear how much code-switching may be occurring mentally as subjects perform tasks. Further, the requirement to respond in a single language will have an impact on speeded tasks that must be considered in interpretation of results. In studies comparing multilingual/bilingual subjects across languages, the focus is generally not on the factors affecting

voluntary choice of terms, but on how a subject's perceptual choice behavior or categorization and naming of a color sample are affected by use of a specific language. Researchers have asked whether perceptual boundaries match the categories existing within a language and whether perceptual experience changes to fit a second language's categories when a new language is acquired. As yet, no studies have investigated whether an individual might prefer to name colors using the language with the richer color vocabulary, the more differentiated category structure, or perhaps the terms most relevant to performing a particular task.

Color Naming

In general, when bilingual individuals perform color naming tasks in two languages, their behavior suggests a dissociation between cognitive processing based on meaning and language choices dictated by grammar or language structure. Thus, a bilingual person who is assimilating to a new culture can make meaning-based choices based on one culture while using language-related naming behaviors typical of the other culture. For example, Pavlenko and Driagina [10] found that English-speaking bilinguals followed a Russian pattern of intransitive verb use when speaking in Russian but switched to an adjectival pattern in English. In contrast, Jameson and Alvarado [4] found that Vietnamese bilinguals tended to impose both the grammar and category structure of English onto Vietnamese responses when responding in Vietnamese. Thus, bilingual subjects responding in Vietnamese tended to name the category orange (a distinct category in English but not Vietnamese) using the object gloss (the fruit name) in both English and Vietnamese, whereas those monolingual in Vietnamese (in both the United States and Vietnam) named the same color sample using a modified basic term (e.g., *Vàng Dam* or dark yellow, with the adjective following the noun). On the other hand, native Vietnamese speakers responding in English were more likely to use Vietnamese language patterns such as more modified basic terms and compound

terms (e.g., yellow-brown) than native speakers of English, who were more likely to use monolexemes (tan for yellowish-brown colors) and object glosses (bark, baby puke). Thus, bilinguals retained the language structure but not the meaning-based categorization patterns [4, 5]. Alvarado and Jameson [5] found a similar dissociation for emotion terms. It is possible this behavior may be specific to the domain of color terms, since Pavlenko [11] has identified several studies in which a domain of terms (in her case, emotion terms) has been treated differently than other categories of concrete and abstract nouns in a language. This suggests the need for greater caution in use of language-based properties such as concreteness/abstraction as a definition for basic color status and consideration of how the content of a domain might interact with language structure. Within-individual comparisons of naming behavior can help identify such differences in language structure as applied to a specific domain of terms, such as color names.

Color surveys that have emphasized monolexemic naming in a search for single-word basic color terms have disadvantaged those languages employing other strategies for making distinctions between colors, including use of modifiers, suffixes and prefixes to stem terms, intensifiers such as word repetition, and use of multiple-character combinations to name basic colors, such as in Asian languages with nonphonetic writing systems. These interactions of grammar with meaning complicate cross-cultural comparisons used to form theories about basic color naming, but can be clarified by examining behavior of bilingual subjects because meaning-based choices are affected differently than grammar-based choices.

Category Boundaries and Focal Colors

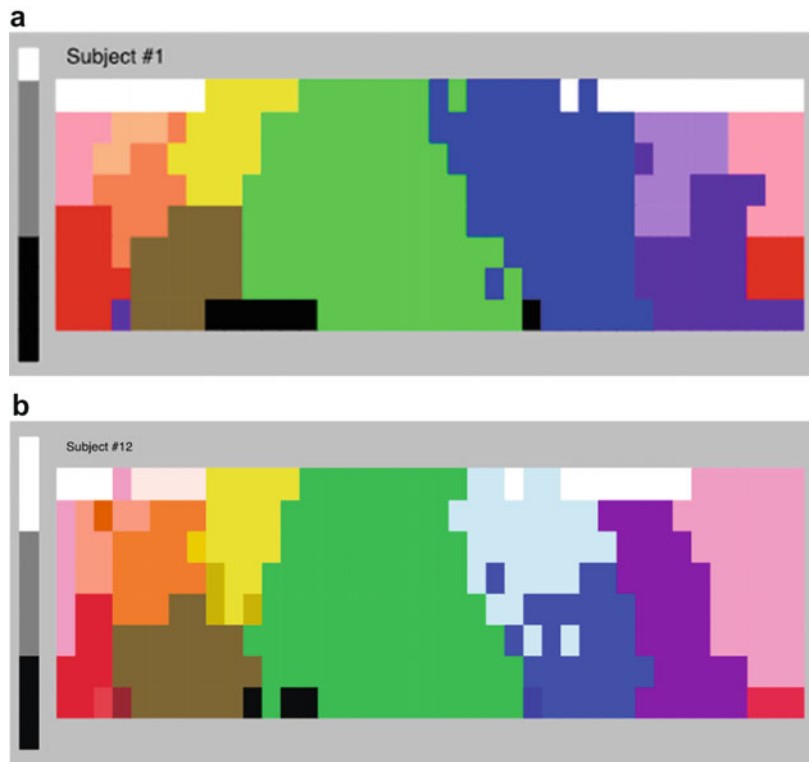
Languages vary in how many color terms they contain and in where they draw category boundaries [1, 2]. Languages in the early stages of **Berlin and Kay's** hierarchy of color category evolution have fewer basic color terms and broader categories with different focal hues. For example, in many Southeast Asian and Asian

languages, it is common to combine the green and blue categories which are distinct in English into a single category (commonly referred to as a “Grue” category). Investigations comparing English speakers with the Tarahumara people of northern Mexico, who also form a broader Grue category [12, 13], suggest that the differences in linguistic distinctions made by the two languages produce differences in decisions about color similarity (see entry on the ► [“Effects of Color Terms on Color Perception and Cognition”](#)). This does not imply that individuals using those languages are unable to recognize differences between green and blue color samples, nor that they cannot perceive distinct hues. As noted above, Asian languages with a Grue category may use a basic color term (such as *Xanh* for Grue in Vietnamese) combined with modifiers to specify the distinct hue (thus *Xanh lá Cây* for leaf green in Vietnamese) of a color sample [3, 4].

In contrast, many languages with more basic color terms than specified by Berlin and Kay's hierarchy have also been described. For example, Russian divides the blue region of color space into two non-overlapping regions, a light/sky blue region and a dark/navy blue region, designated by the basic color terms *goluboj* and *sinij*, respectively [14]. Influences of this category distinction were investigated in visual search tasks by Winawer et al. [15], who found that Russian speakers were able to discriminate two colors faster when they fell into different blue categories in Russian, and this advantage was not shown by English speakers who do not make this category distinction (see entry on ► [“Color Categorical Perception”](#)). Similar distinctions in the blue region of color space are made in Greek [16], Japanese [17], Italian [18], and Ukrainian [19]. Figure 1 provides a sample comparison of color naming by an English-speaking subject [20] and a Japanese-speaking subject [21], differing in the basic color terms used to name blue color samples. Studies have compared monolingual speakers of these languages to bilingual speakers, where in all cases, English was the second language learned. These studies find bilinguals' color categories to be shifted toward those of monolingual speakers of English across a range of tasks,

Bilingual Color Naming/ Categories, Fig. 1

Monolexemic color naming of 330 Munsell color samples, used in the World Color Survey. (a) English-speaking subject #1 [from 20], (b) Japanese-speaking subject #12 [from 21]. As illustrated by these individual subjects, English speakers use one basic color term to name blue color samples, while Japanese speakers use two basic color terms, *mizu* to name light blue color samples and *ao* to name medium to dark blue color samples



including color similarity judgments and visual search (through which categorical effects for color perception have also been demonstrated) [22, 23] and even verbal descriptions of visually more complex stimuli [24]. Paramei et al. [25] however find a shift in the opposite direction, with the majority of their Italian-English bilinguals selecting the foci of English blue to be closer to the foci of Italian *blu* (dark blue). Extralinguistic variables such as second language proficiency, frequency of second language use, and length of stay in second language-speaking country have been suggested to influence category representations of the bilinguals in all of these studies. Additionally, as the last study suggests, linguistic variables such as the phonological and orthographic similarity of color terms in the languages in question (such as with the color terms for English blue and Italian *blu*) may also play a role.

While nearly all of the work discussed above has looked at behavioral evidence for language effects on color, Athanasopoulos et al. [22] find converging neuro-physiological evidence, focusing on an ERP component indicating pre-attentive

differences in the perception of color in monolingual and bilingual speakers (again, see entry ► “Color Categorical Perception”). As has been pursued in other domains of color research (see entries ► “Bayesian Approaches to Color Category Learning” and ► “Color Category Learning in Naming-Game Simulations”), computational models are now being explored to explain behavioral findings in multilingual/bilingual color naming. Matussevych et al. [26], for example, model the shift in bilinguals’ category foci toward the foci of the second learned language, examining the influence of category similarity across languages, and translations for color terms being shared (or not) across languages. As noted by them, their modeling framework could be expanded to more systematically test the influences of the various other linguistic and extralinguistic variables noted previously. Other approaches such as connectionist modeling are also well suited to examining the representation of color categories in the mind of multilingual/bilingual individuals. Zhang et al. [27], for example, present a connectionist model of object

naming, but applications to color naming have not yet been explored.

Color Lexicon Evolution

Since the specification of **Berlin and Kay's Theory** of the evolution of color naming systems, several modifications have been proposed. MacLaury [28] suggests that Berlin and Kay's evolutionary stages reflect a transition from a lightness-based naming system to a hue-based system and that cultures evolve in their naming due to a shift in emphasis from describing similarity to describing difference (see entry on ► ["Vantage Theory of Color"](#)). Schirillo [29] suggests that evolutionary stages may also reflect a transition from use of contextualized names to use of more abstract terms (basic terms are abstract by definition). Van Brakel [30] suggests that stages reflect the influence of Western culture on indigenous naming behaviors and are thus a transition to the Western color system from a variety of viable alternatives (see entry on ► ["Color Dictionaries and Corpora"](#)). However, in studies of unconstrained naming, where subjects are not confined to use of basic color terms, both English speakers and those speaking other languages greatly prefer to describe colors using secondary names or modified basic terms, not basic terms alone. Table 1 provides an example of the variety of basic and non-basic secondary color terms, compound words, and modifiers, used by English and Korean monolinguals, and Korean/English bilinguals, in naming of a color sample (C23 from the World Color Survey palette) [1, 9]. Another example, of British English and Mandarin Chinese color naming, can be found in the work of Lin et al. [31]. If cultures were evolving toward a greater emphasis on describing difference, those cultures with more basic terms should also have more modifiers. The opposite seems to be true – use of modifiers and use of contextualized basic terms seem to vary inversely, suggesting that these are alternative strategies for describing differences in color percepts.

The behavior of multilingual/bilingual speakers may recapitulate this evolutionary process on the individual level to the extent that

immigrants speaking multiple languages are in transition to English from one or more other languages. It can be argued that generalizing from individual language learning to evolution of language stages at the cultural level is justified because (a) the culture's language practices are determined by the behavior of individuals speaking that language; (b) as with individuals, cultures change in part due to contact with other cultures; and (c) the culture's evolutionary stage is measured by assessing the language behaviors of participants sampled from that culture. Otherwise, the processes by which cultures change are not hypothesized to be the same as those by which individuals acquire language abilities. In support of this idea, Jameson and Alvarado [4] found that Vietnamese bilinguals used a basic color term for orange while Vietnamese monolinguals did not and that the focal samples for bilinguals responding in Vietnamese had shifted toward those of monolingual English speakers. However, for these Vietnamese bilinguals, highly specific Vietnamese object glosses and modifiers tended to fall into disuse before basic color terms. Thus, bilingual speakers were unable to fully emulate the pattern of naming used by monolingual English speakers because they did not have access to a large vocabulary of object glosses in Vietnamese (probably due to disuse). More recently, Pavlenko et al. [24] found that Russian monolinguals made much finer distinctions than English monolinguals and bilinguals in describing the color content of paintings such as *The Blue Window* by Matisse and *The Starry Night* by Van Gogh, containing rich variations in shades of blue. As previously noted, Russian has two basic color terms for light and dark blue, while English has one basic color term for the blue region of color space. Just as speakers of Grue languages use modifiers to describe the differences they perceive between green and blue color samples, English speakers can certainly perceive differences in shades of blue and would be able to describe the differences using modifiers such as *Light* and *Sky* or secondary color terms such as *aqua* and *turquoise*, if prompted to do so. However, this behavior may not be observed spontaneously because the distinctions between shades of blue do not hold communicative

Bilingual Color Naming/Categories, Table 1 Unconstrained color naming of a color sample (C23 from the World Color Survey palette) by three language groups:

(1) English monolinguals, (2) Korean/English bilinguals naming in both English and Korean, and (3) Korean monolinguals [9]

English monolinguals (<i>n</i> = 32)		Bilinguals naming in English (<i>n</i> = 34)		Bilinguals naming in Korean (<i>n</i> = 34)		Korean monolinguals (<i>n</i> = 11)	
Green	22%	Mint	35%	Minteu-saek (mint)	24%	Minteu-saek (mint)	36%
Seafoam green	16%	Light mint	9%	Chorok-saek (grass green)	21%	Balg-eun minteu-saek (bright mint)	18%
Teal	13%	Light green	9%	Emelaldeu-saek (emerald)	12%	Yen nok-saek (light green)	9%
Mint	3%	Green	6%	Phalan/phalangsae (blue/green)	9%	Balg-eun yentwu-saek (bright yellow green)	9%
Seafoam	3%	Emerald	6%	Yentwu-saek (yellow green)	6%	Emelaldeu nok-saek (emerald green)	9%
Pastel green	3%	Teal	6%	Ogsaek (jade)	3%	Pichwi-saek (jade green)	9%
Pond green	3%	Turquoise	6%	Turquoise	3%	Yeonhan haneul-saek (light sky blue)	9%
Light blue green	3%	Mint green	3%	Pichwi-saek (jade green)	3%		
Blue	3%	Blue	3%	Balg-eun nok-saek (bright green)	3%		
Light blue	3%	Light blue	3%	Yeonhan chorok-saek (light green)	3%		
Barely blue	3%	Light sky blue	3%	Haneul-saek (sky blue)	3%		
Aqua	3%	Pastel blue	3%	Yethun haneul-saek (light sky blue)	3%		
Turquoise	3%	Cyan	3%	Baby pullwu-saek (baby blue)	3%		
Light turquoise	3%	Aquamarine	3%	Minteu-saek saljjak yeonhan (mint a bit light)	3%		
Pale yellow	3%	Yellowish green	3%				

relevance in their language community. The same explanation may hold for the bilingual speakers, who were either from Russian-speaking families but brought up in an English language environment in the United States or were new learners of Russian in the United States. For second language learners of Russian in a Russian language environment, the light/dark blue distinction may be expected to become more salient, as the distinction would be obligatorily made by native Russian speakers. A comparison of multicultural multilingual naming with immigrant multilingual naming is needed to show whether those with richer language experience show an increased tendency to

use their wider vocabularies to describe additional dimensions of perceptual experience.

Alvarado and Jameson [5] suggest that cultures may not evolve from lightness to hue but rather from a single-dimensional naming system (based on lightness) to a multidimensional naming system capturing distinctions related to hue and saturation in addition to lightness. If so, the interaction between dimensions may produce more sharply defined category boundaries than a single dimension alone. In that case, modifier use should increase as cultures attend to additional dimensions. Further, modifiers tend to evolve along with basic terms. As noted, Tarahumara terms like Very and Somewhat evolve to Light

and Bright, terms which are specific to color naming. Vietnamese uses the term Fresh (*Tươi*) to describe saturation, a modifier not used in English but equivalent to terms such as Vivid or Deep. Areas of color space described by lightness modifiers do not vary as much as those described by modifiers capturing other dimensions. This suggests that an evolutionary-stage theory based on a broader spectrum of naming behavior, capturing more aspects of spontaneous naming behavior, might produce a different ordering of stages.

Theoretical Considerations

Comparisons of naming behavior of multilingual individuals across languages provide an approach to testing assumptions of current theories of color naming, such as the assumption that monolexemic naming using basic color terms is a valid approach to comparing languages with widely varying strategies for describing color appearances. For example, as described by Burgess et al. [12], the Tarahumara language used a postposed bound modifier (a modifier added to the end of a stem term) that specifies the relation of the currently named color to the center of the category. This system is different from Chinese, Vietnamese, and English [31], where modifiers are separate words and relational distinctions are subordinated to other distinctions. In contrast, Chinese characters are frequently constructed by compounding several characters into a single character with a new meaning, and all color terms consist of two-character pairs in which the first character specifies the color and the second character specifies that it is a noun referring to the color appearance itself (decontextualized, abstract). When the color name is used as an adjective, the second character is the name of the object taking that color. In Japanese, a different second character is used to differentiate chromatic and achromatic color names [31]. This complicates comparison of the basicness of terms when the focus is on monolexemic naming and on the properties of the language instead of the behavior of the individuals speaking that language. Emphasis on a few basic color categories, in both study and theory, neglects the richness of color naming behavior.

It has been argued that immigrant monolingual speakers in the United States (or elsewhere) are not a suitable population for cross-cultural investigations. When monolingual speakers in immigrant communities do not acquire a new language, there appears to be a close similarity between their responses and those of monolingual speakers in their original country of residence. This may be most true for immigrants living in large insulated neighborhoods in urban areas who can conduct daily business without needing to speak an additional language. In two studies of different domains [4, 8], monolingual Vietnamese speakers in the United States and Saigon produced responses that were closely similar, suggesting that the groups responding in the United States were a close proxy for their counterparts in Vietnam. Cultural contamination cannot be automatically presumed to invalidate every investigation. Although it might also be argued that the availability of Western media may have changed the Vietnamese lexicon, resulting in the close similarity of results among the studies, differences were noted for certain contextualized terms which parallel the differences observed between the American and Vietnamese cultures themselves. For example, differences in categorization and naming of the category orange were observed for bilingual subjects but not for monolingual immigrants living in the United States, who would presumably have been exposed to Western media. Given that such cultural differences were not eliminated by exposure to the media, perhaps everyday social interaction and communication have a greater impact on language than passive media exposure. Findings of the persistence of regional accents in Great Britain in the face of homogenous media speech support this conclusion. To assess this, the level of assimilation should be measured in studies of multilingual/bilingual versus monolingual immigrant speakers.

Study of the cognitive language processing of multilingual/bilingual individuals in color naming tasks has the potential to widen our understanding of how language is used to describe perceptual experience by providing a glimpse into the cognitive processing that can be attributed to meaning distinct from the processing that is related to language, holding individual differences and cultural

differences constant. However, to conduct such studies, comparisons must be broadened to compare bilingual and monolingual subjects in both the native country and the country to which such subjects have emigrated, as well as comparisons between multilingual individuals within multicultural societies in which multiple languages are regularly spoken independent of immigrant experience.

Cross-References

- [Bayesian Approaches to Color Category Learning](#)
- [Berlin and Kay Theory](#)
- [Color Categorical Perception](#)
- [Color Category Learning in Naming-Game Simulations](#)
- [Color Dictionaries and Corpora](#)
- [Effects of Color Terms on Color Perception and Cognition](#)
- [Memory Color](#)
- [Vantage Theory of Color](#)
- [World Color Survey](#)

References

1. Kay, P., Berlin, B., Maffi, L., Merrifield, W., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
2. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
3. Jameson, K.A., Alvarado, N.: Differences in color naming and color salience in Vietnamese and English. *Color. Res. Appl.* **28**, 113–138 (2003)
4. Alvarado, N., Jameson, K.A.: The use of modifying terms in the naming and categorization of color appearances in Vietnamese and English. *J. Cogn. Cult.* **2**(1), 53–80 (2002)
5. Alvarado, N., Jameson, K.A.: Shared knowledge about emotion among Vietnamese and English bilingual and monolingual speakers. *J. Cross-Cult. Psychol.* **42**, 964–983 (2011)
6. Loitšenko, O.: Colour terms in the BLUE area among Estonian-Russian and Russian-Estonian bilinguals. In: *Progress in Colour Studies: Cognition, Language and Beyond*. John Benjamins Publishing Company, Amsterdam (2018)
7. Athanasopoulos, P.: Cognitive representation of colour in bilinguals: the case of Greek blues. *Biling. Lang. Cogn.* **12**(1), 83–95 (2009)
8. Bullock, B.E., Toribio, A.J. (eds.): *The Cambridge Handbook of Linguistic Code-Switching*, 1st edn. Cambridge University Press, London (2009)
9. Deshpande, P.S.: *Color categorization in bilingual population: Korean-English*. University of California, Irvine, Undergraduate thesis (2015)
10. Pavlenko, A., Driagina, V.: Russian emotion vocabulary in American learners' narratives. *Mod. Lang. J.* **91**, 213–234 (2007)
11. Pavlenko, A.: Emotion and emotion-laden words in the bilingual lexicon. *Biling. Lang. Cogn.* **11**, 147–164 (2008)
12. Burgess, D., Kempton, W., MacLaury, R.E.: Tarahumara color modifiers: category structure presaging evolutionary change. *Am. Ethnol.* **10**(1), 133–149 (1983)
13. Kay, P., Kempton, W.: What is the Sapir-Whorf hypothesis? *Am. Anthropol.* **86**(1), 65–79 (1984)
14. Davies, I., Corbett, G.: The basic color terms of Russian. *Linguistics*. **32**(1), 65–90 (1984)
15. Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., Boroditsky, L.: Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci.* **104**(19), 7780–7785 (2007)
16. Androulaki, A., Pestana, N.G., Lillo, J., Davies, I.R.: Greek colour terms: evidence for a natural category boundary within blue. *Percept. ECVF Abstr.* **30**, 0–0 (2001)
17. Uchikawa, K., Boynton, R.M.: Categorical color perception of Japanese observers: comparison with that of Americans. *Vis. Res.* **27**, 1825–1833 (1987)
18. Paramei, G.V., D'Orsi, M., Menegaz, G.: 'Italian blues': a challenge to the universal inventory of basic colour terms. *J. Int. Colour Assoc.* **13**, 27–35 (2014)
19. Starko, V.: Ukrainian colour concepts for blue. *J. Slavic Lang. Lit.* **54**, 150–163 (2013)
20. Lindsey, D.T., Brown, A.M.: The color lexicon of American English. *J. Vis.* **14**(2), 17–17 (2014)
21. Kuriki, I., Lange, R., Muto, Y., Brown, A. M., Fukuda, K., Tokunaga, R., . . . Shioiri, S.: The modern Japanese color lexicon. *J. Vis.* **17**(3), 1–1 (2017)
22. Athanasopoulos, P., Dering, B., Wiggert, A., Kuipers, J.R., Thierry, G.: Perceptual shift in bilingualism: brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*. **116**(3), 437–443 (2010)
23. Athanasopoulos, P., Damjanovic, L., Krajciová, A., Sasaki, M.: Representation of colour concepts in bilingual cognition: the case of Japanese blues. *Biling. Lang. Cogn.* **14**(1), 9–17 (2011)
24. Pavlenko, A., Jarvis, S., Melnyk, S., Sorokina, A.: Communicative relevance: color references in bilingual and trilingual speakers. *Biling. Lang. Cogn.* **20**(4), 853–866 (2017)
25. Paramei, G.V., D'Orsi, M., Menegaz, G.: Cross-linguistic similarity affects L2 cognate representation: blu vs. blue in Italian-English bilinguals. *J. Int. Colour Assoc.* **16**, 69–81 (2016)
26. Matusevych, Y., Beekhuizen, B., Stevenson, S.: Crosslinguistic transfer as category adjustment: modeling conceptual color shift in bilingualism. In: Kalish, C., Rau, M., Rogers, T., Zhu, J. (eds.)

- Proceedings of the 40th Annual Conference of the Cognitive Science Society, pp. 744–749. Cognitive Science Society, Austin (2018)
27. Fang, S.Y., Zinszer, B.D., Malt, B.C., Li, P.: Bilingual object naming: a connectionist model. *Front. Psychol.* **7**, 644 (2016)
 28. MacLaury, R.: From brightness to hue: an explanatory model of color-category evolution. *Curr. Anthropol.* **33**, 137–163 (1992)
 29. Schirillo, J.: Tutorial on the importance of color in language and culture. *Color. Res. Appl.* **26**, 179–192 (2001)
 30. Van Brakel, J.: Comment on “from brightness to hue”. *Curr. Anthropol.* **33**, 169–172 (1992)
 31. Lin, H., Luo, M., MacDonald, L., Tarrant, A.: A cross-cultural colour-naming study. Part I: using an unconstrained method. *Color. Res. Appl.* **26**, 40–60 (2001)

Billmeyer, Fred Wallace, Jr.

Ellen C. Carter

Color Research and Application and, Konica
Minolta, Pennsville, NJ, USA



Fred Wallace Billmeyer, Jr., was an American chemist who contributed through industrial

research and academia to the developing fields of polymer chemistry and color science during the second half of the twentieth century.

Billmeyer was born on August 24, 1919, in Chattanooga, Tennessee. In 1941, he received a B.S. in chemistry from the California Institute of Technology. Then, he received a Ph.D. in physical chemistry from Cornell University in 1945 after studying the measurement of molecular weights by light scattering under Nobel Laureate Peter Debye.

His professional career can be divided into two major parts, focusing first on industry and then on education. Upon graduation from Cornell, he joined the Plastics Department of E. I. du Pont de Nemours and Company, Wilmington, Delaware, where he remained for 20 years. There he worked on various aspects of polymers. Instinctively an educator, he published *Textbook of Polymer Chemistry* (1957) and *Textbook of Polymer Science* (1962), which went through three editions and was also published in Japanese. These books helped to create the academic discipline of polymer science. He served as a visiting professor in chemical engineering at Massachusetts Institute of Technology for the 1960–1961 academic year. Billmeyer increasingly became involved with the growing field of color science. At the University of Delaware, he supervised his first graduate color science student, Joseph Atkins, whose 1964 Ph.D. thesis was on the Absorption and Scattering of Light in Turbid Media.

In 1964, he retired from DuPont to join academia full time as a professor of analytical chemistry at Rensselaer Polytechnic Institute (RPI), Troy, New York. At RPI, he taught and directed research in both polymer and color science and directed the Rensselaer Color Measurement Laboratory until he became professor emeritus in 1984. During his time at RPI, he not only taught and mentored undergraduate and graduate students but also, with Max Saltzman, initiated summer courses about color science for people from industry. His influence on the field of color science was extended throughout the United States by over 1000 people from diverse fields attending these summer programs.

Deepening his commitment to education, he undertook two other major projects while at RPI. In 1966, Professor Billmeyer continued his collaboration with Max Saltzman, publishing *Principles of Color Technology*. Their two editions of this textbook became widely acclaimed throughout the industrial world as the primary introduction to color science. A third edition entitled *Billmeyer and Saltzman's Principles of Color Technology* (published in 2000) by Roy Berns is still widely used. Also with Richard Kelly in 1975, he published *Entering Industry: A Guide for Young Professionals*.

The second project was the creation of an academic journal for color science. After supporting two unsuccessful attempts at journals of color engineering and science, in 1976, he initiated the journal *Color Research and Application*. Billmeyer was the founding editor and served in that position until 1986. His approach of involving national color organizations, beginning with the Canadian Society for Color, the US Inter-Society Color Council, and the Colour Group (Great Britain) and later others throughout the world, has contributed to the success of this English-language journal published by John Wiley & Sons now in its 40th year of continuous publication.

Many of Billmeyer's scientific contributions throughout the years fall into the category of light scattering and the application of turbid medium theory to the formulation and shading of colored materials. Beginning with the basic theory of Beer's law, he programmed a digital computer to calculate the concentrations required to produce a given transparent color by mixing soluble dyes [1]. Next with Beasley and Atkins, he expanded two-flux Kubelka-Munk theory to a four-flux turbid medium theory [2], which became known as the BAB theory for colorant formulation. From there, he continued the development of the BAB to apply particularly to paint systems [3] and expanded that further to coatings with goniometric characteristics [4, 5]. He further developed practices for determining the optical properties needed for use in the turbid medium theory and multi-flux theory [6–9].

Billmeyer was a versatile and prolific communicator; his contributions were not limited to turbid medium theory. Two other areas that should be mentioned here are the characterization of fluorescent materials [10, 11] and use of color difference equations in manufacturing [12]. His involvement within the many professional organizations and collaboration with other scientists throughout the world influenced most areas of color science. In particular, his more than 350 published articles discussed polymer characterization and diverse aspects of color. His 13 books presented in clear, concise language the scientific principles necessary for the industrial use of polymers and color to enhance our lives.

Major Books

- Billmeyer, F.W.: Textbook of Polymer Chemistry. Interscience, New York (1957) (in Russian 1958)
- Billmeyer, F. W.: Textbook of Polymer Science, 3 edn. Wiley, New York (1962, 1971, 1884) (in Japanese 1969 & 1989; in Malaysian 1989)
- Billmeyer, F. W., Saltzman, M.: Principles of Color Technology, 2 edn. Wiley, New York (1966, 1981) (in German Grundlagen der Farbtechnologie. Goettingen: Zeurich: Muster-Schmidt Verlag, 1993)
- Billmeyer, F.W.: Synthetic Polymers. Double Day, Garden City (1972)
- Collins, E.A., Bares, J., Billmeyer, F.W.: Experiments in Polymer Science. Wiley-Interscience, New York (1973)
- Billmeyer, F.W., Kelley, R.N.: Entering Industry: A Guide for Young Professionals. Wiley, New York (1975)

References

1. Billmeyer, F.W., Beasley, J.K., Sheldon, J.A.: Formulation of transparent colors with a digital computer. *J. Opt. Soc. Am.* **50**, 70–72 (1960)
2. Beasley, J.K., Atkins, J.T., Billmeyer, F.W.: Scattering and absorption of light in turbid media electromagnetic scattering. In: Proceedings of the 2nd Interdisciplinary Conference on Electromagnetic Scattering, pp. 765–785. University of Massachusetts, Amherst (1965)
3. Billmeyer, F.W., Abrams, R.L.: Predicting reflectance and color of paint films by Kubelka-Munk analysis i – turbid medium theory. *JPT.* **45**(579), 23–38 (1973)
4. Billmeyer, F.W., Davidson, J.G.: Color and appearance of metalized paint films I. Characterization. *J. Paint Tech.* **46**(593), 31–37 (1974)

5. Billmeyer, F.W., Carter, E.C.: Color and appearance of metalized paint films II. Initial application of turbid-medium theory. *J. Paint Tech.* **48**(613), 53–60 (1976)
6. Billmeyer, F.W., Chassaigne, P.G., Dubois, J.F.: Determining pigment optical properties for use in the Mie and many-flux theories. *Color Res. Appl.* **5**, 108–112 (1980)
7. Billmeyer, F.W., Phillips, D.G.: Predicting reflectance and color of paint films by Kubelka-Munk analysis III effect of concentration errors on color for mixtures of one chromatic pigment with white. *J. Paint Tech.* **46**(592), 26–29 (1974)
8. Billmeyer, F.W., Phillips, D.G.: Predicting reflectance and color of paint films by Kubelka-Munk analysis IV Kubelka Munk scattering coefficient. *J. Paint Tech.* **48**(616), 30–36 (1976)
9. Billmeyer, F.W., Richards, L.W.: Scattering and absorption of radiation by lighting materials. *J. Color Appearance.* **2**, 4–15 (1973)
10. Alman, D.H., Billmeyer, F.W.: Integrating-sphere errors in the colorimetry of fluorescent materials. *Color Res. Appl.* **1**, 114–145 (1976)
11. Alman, D.H., Billmeyer, F.W.: New method of colorimetric measurement of opaque fluorescent samples. *Color Res. Appl.* **2**, 19–25 (1977)
12. Morley, D.I., Munn Rich, R., Billmeyer, F.W.: Small and moderate colour differences parts I and II. *J. Soc. Dyers Colour.* **91**(229–242), 302–307 (1975)

Binocular Color Matching

David R. Simmons
School of Psychology, University of Glasgow,
Glasgow, Scotland, UK

Synonyms

Binocular color perception; Binocular color vision

Definition

In the context of this article, “binocular color matching” refers to color vision with two eyes, focusing on the differences between when two eyes are used to view colors (i.e., binocular or dichoptic color) and when only one is used (i.e., monocular color), as well as phenomena

(largely) unique to binocular vision such as fusion, rivalry, and stereopsis.

Introduction

Most of us have two functioning eyes, and yet we still somehow see the visual world as a single entity, most of the time unaware that our brains are receiving and processing two separate and largely independent sets of visual signals from the outside world. In the context of color vision, this raises some interesting questions:

1. *Binocular color fusion*: How similar do the chromatic properties of visual signals in each eye have to be in order to be perceived as one?
2. *Binocular color rivalry*: How different do the chromatic properties of visual signals in each eye have to be in order to evoke binocular rivalry, which is an alternation in visibility of the visual information going to each eye?
3. *Color and stereopsis*: Stereopsis is the perception of depth that is obtained by analyzing the differences between the images going to each eye. To what extent is this perception of depth affected by the chromatic properties of visual information?
4. *Binocular color summation*: The processing of visual signals is affected by whether or not a scene is viewed with one or two eyes: to what extent do the chromatic properties of visual signals influence this?
5. *Binocular color appearance*: Do colors look different when viewed with one eye or two?

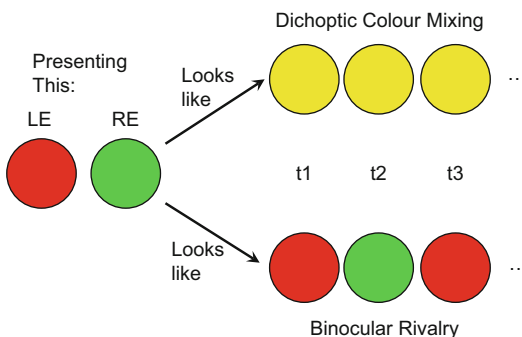
Note that the field of binocular vision and stereopsis has been thoroughly reviewed by Prof. Brian Rogers of the University of Oxford, UK, and Prof. Ian Howard, late of York University, Canada [11], and for all but the most recent studies, the interested reader should refer to them for further information.

Binocular Color Fusion

The problem of binocular fusion, how the world is seen as one despite having two eyes, has fascinated

philosophers and scientists for centuries (see [11] for a historical overview). It makes sense that color might have a role to play in this phenomenon. One might expect a feature in the visual world when viewed with one eye to have the same spectral reflectance properties and therefore presumably appear to be the same color, when viewed with the other eye. Consequently, interocular color similarity would presumably make a good visual cue for fusion. Curiously, however, most of the research effort has gone into investigating what happens when the colors presented to the two eyes are *different*. In particular, there is the controversial issue of *dichoptic color mixture*: when two lights with different chromatic contents are presented separately to each eye, do the lights combine in a similar way to when they are presented superimposed to the same eye or does *binocular rivalry* (an alternation in perception between the stimuli presented to each eye) result (see Fig. 1)? The first investigation of this phenomenon appears to have been in the early eighteenth century using differently colored silks viewed through an aperture [7]. Later, in the nineteenth century, the existence or not of dichoptic color mixture was a further source of controversy between Helmholtz and Hering, given the different predictions of the Young-Helmholtz and Hering theories of color vision. Arguably the most controversial aspect is whether or not “binocular yellow” can be obtained when red light is presented to one eye and green to the other. Recent studies have clarified that true dichoptic color mixture does appear to take place, so long as certain

conditions are met. These are that the mixture is more stable with small and textured patches of light rather than large and homogeneous, with flickering rather than steady stimuli, and with patches of low luminance and saturation which are close in luminance and chromaticity rather than high luminance and saturation with different luminances and chromaticity ([11]; see also [10]). Some of these rules are nicely illustrated in a study by de Weert and Wade [5] (see Fig. 2). In a more recent study, the bold claim is made that “two eyes are worse than one,” based on data suggesting that monocularly visible features of different colors effectively disappear when dichoptically combined [1]. In an earlier study by Simmons [32], however, although detection thresholds for briefly presented iso-luminant red-green gratings in antiphase between the eyes were higher than those for monocular presentation, they were not so high that the signals from each eye were effectively canceling each other out, as suggested by Anstis and Rogers [1]. Potentially, technical issues to do with the accurate registration of stimuli in each eye and the presence



Binocular Color Matching, Fig. 1 Schematic diagram illustrating alternative percepts generated by dichoptic color mixing or binocular rivalry



Binocular Color Matching, Fig. 2 Dichoptic color mixing: fusion of the *upper* two disks gives unstable binocular rivalry; fusion of the *lower* two textured disks produces stable dichoptic color mixing. This illustrates the role of texture in promoting dichoptic color mixing (Taken from de Weert and Wade [5], Reprinted from Vision Research, 28, Charles M.M. de Weert, Nicholas J. Wade, Compound binocular rivalry, 1031–1040, Copyright (1988), with permission from Elsevier)

of barely detectable luminance artifacts complicate the interpretation of these experiments.

A different approach was taken by Malkoc and Kingdom [21], who, following on from an earlier study by Yoonessi and Kingdom [42], measured *dichoptic color difference thresholds (DCDTs)*, which are the thresholds for detecting a color difference between the two eyes, coinciding with the detection of a peculiar phenomenon called *binocular luster*. They found that these thresholds were higher than those for detecting color differences between two stimuli when they were presented side by side, although lower than those required to provoke binocular rivalry (see below). Malkoc and Kingdom [21] also found that these thresholds were best predicted by the perceived color difference between the two stimuli, rather than any considerations based on cardinal or unique hue mechanisms. Kingdom and Libenson [14] specifically investigated the processing of interocular differences in saturation (or color contrast). They found that the appearance of the mixture obtained crucially depended on the relative amounts of luminance and chromatic contrast. With purely chromatic differences between the two eyes (i.e., lights with the same hue but different saturations), the more saturated/higher contrast stimulus dominated the percept, but the presence of a luminance pedestal forced the colors to blend and therefore reduced the saturation of the resultant. Kingdom and Libenson [14] argue from these results that the appearance of a dichoptic color mixture depends on whether or not the brain interprets the information from the two eyes as coming from the same object or not, a phenomenon which they term the “object commonality hypothesis.”

Binocular Color Rivalry

As rivalry is, to a certain extent, the obverse of fusion, most of the issues pertinent to this theme have already been discussed above. It is certainly true that the presentation of saturated red stimuli to one eye and saturated green to the other is often almost paradigmatic in studies where the aim is to evoke binocular rivalry (see, e.g., [2]; also, again,

[5]). O’Shea and Williams [23] demonstrated that S-cone-isolating stimuli could induce binocular rivalry, suggesting that rivalry was not solely confined to luminance or red-green chromatic pathways. A detailed study of the wavelength sensitivity of binocular rivalry was performed by Sagawa [28]. Note that rivalry does not occur for briefly presented stimuli. In these situations the dichoptic stimuli tend to superimpose, although are still distinguishable from monocularly superimposed stimuli [11]. There is some evidence that the chromatic system is more affected by binocular rivalry suppression than the achromatic system [11, 24]. Mullen et al. [22] found that the visibility of chromatic grating stimuli presented to one eye was affected by the presence of luminance stimuli in the other, suggesting that, when stimuli differ between the two eyes, the suppression of one eye by the other is independent of whether the stimulus contains color or luminance contrast (which is not the case for monocular vision, or when the stimulus is the same in both eyes, when the suppression of one stimulus by another, masking, stimulus is more selective).

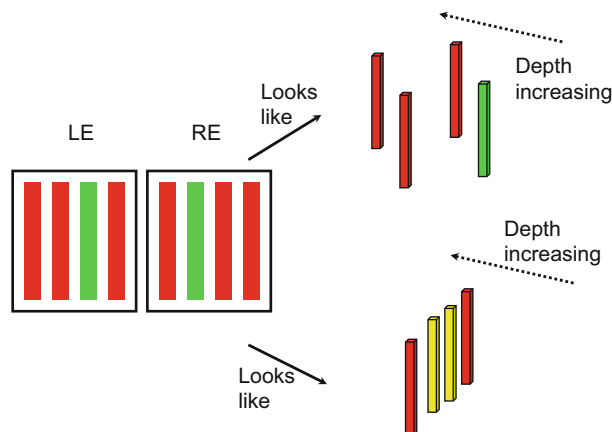
Color and Stereopsis

There are two issues which have dominated research on color and stereopsis. The first is whether or not there is actually a functioning chromatic stereopsis mechanism. The second is whether or not chromatic information helps to solve the stereo correspondence problem. In both cases the arguments have been similar to those in color motion perception. Historically, a good way of testing whether or not a purely chromatic mechanism exists has been to test performance at *isoluminance*, when the visual patterns in question are theoretically defined solely by chromatic contrast, without any luminance contrast being present. The earliest study of stereopsis at isoluminance suggested that stereoscopic depth perception was very weak and possibly nonexistent at red-green isoluminance [20]. Subsequent studies provided conflicting results, depending on the precise stimulus characteristics [11]. Around the early 1990s, the extreme views on this issue

were represented by two studies: Livingstone and Hubel [19] and Scharff and Geisler [29]. Livingstone and Hubel [19] claimed that all previous demonstrations of intact stereoscopic depth perception were artifactual, and down to the technical difficulties associated with removing all luminance information from the stimulus. Their view that stereopsis was essentially “color blind” was consistent with their theories on the parallel processing of visual information by magnocellular- and parvocellular-mediated visual pathways, with stereopsis mediated by the “color-blind” magnocellular stream. Scharff and Geisler [29], on the other hand, claimed that their careful calibration of the stimulus, accounting for all potential luminance artifacts, not only demonstrated that stereopsis was possible at isoluminance but that it was as good as it could be, given the limits on obtainable color contrast due to the overlap in spectral sensitivities of long- and medium-wavelength-sensitive cones. In a series of studies published between 1994 and 2002, Simmons and Kingdom updated this view of the status of stereopsis at isoluminance to demonstrate that there exists a rudimentary chromatic stereopsis mechanism which is less contrast sensitive and has a more limited disparity range, poorer stereoacuity, and poorer ability to encode a stereoscopically

defined shape than its luminance counterpart ([15, 17, 33–35]; reviewed in [16]). While this view was challenged in a study by Krauskopf and Forte [43], who argued, similarly to Livingstone and Hubel [19], for a complete disappearance of stereopsis at isoluminance, their data could be partially explained by the presence of high spatial frequency luminance artifacts in their stimuli. Having said that, it is very difficult to control for these artifacts, especially given the high contrast sensitivity of luminance-based stereopsis [31, 37]. While most of the above studies were carried out at red-green isoluminance, Grinberg and Williams [9] demonstrated that stereopsis is also possible at “blue-yellow” isoluminance.

Can chromatic information help to solve the stereo correspondence problem? The correspondence problem in stereopsis refers to the ability of the human brain to work out which feature in one eye’s view matches with which feature in the other eye’s view (see Fig. 3). As stated above, it makes sense that chromatic information could help in solving this problem, and, sure enough, a number of studies have demonstrated that this is indeed the case (e.g., [6, 12, 18]). Simmons and Kingdom [36] argued that these results might be explicable in terms of the interactions between different



Binocular Color Matching, Fig. 3 Schematic illustration of the potential role of color in the stereo correspondence problem. If the images illustrated are presented to the left and right eyes, respectively, if color promotes stereo correspondence, then a *green bar* will appear in front and a

red bar behind the fixation plane (*upper right*). If color has no role, then four bars will be seen in the frontoparallel plane (*lower right*). The bars colored *yellow* might also appear rivalrous, depending on the precise stimulus conditions

independent stereopsis mechanisms, some of which are sensitive to chromatic contrast (see also [38]). Wardle and Gillam [41] have argued that color information is also important in the so-called da Vinci stereopsis, which is the perception of depth obtained, under certain viewing conditions, from visual regions only visible to one eye.

There are some complicated dependences of stereoscopic sensitivity on stimulus dynamics at isoluminance which were investigated by Tyler and Cavanagh [39]. Note that the phenomenon of *chromostereopsis*, most often experienced as a saturated red target appearing in front of a saturated blue one (when viewed with two eyes) despite the stimuli being physically in the same depth plane, is thought to be largely due to chromatic aberrations in the eye inducing a relative difference (i.e., disparity) in the retinal locations of the red and blue objects in each eye [11, 25]. Figure 4, especially if presented on a data projector and viewed in a dark room, normally induces a strong sensation of this illusory depth percept.

Binocular Color Summation

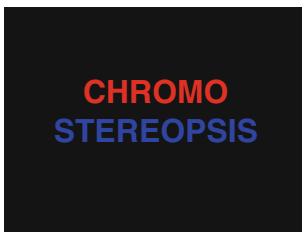
To what extent does viewing a chromatic stimulus with two eyes affect its visibility or the performance obtained with it? The scientific literature presents a somewhat confusing picture, with Anstis and Rogers [1] claiming that “two eyes are worse than one,” Tyler and Cavanagh [39] claiming that “two eyes” are “as sensitive as one,” and Simmons [32] finding evidence for

binocular contrast summation close to “full summation,” when contrast is effectively linearly added between the eyes, and so a binocular stimulus is twice as detectable as when presented to both eyes rather than one (this would be “two eyes are twice as good as one”!). The conditions under which dichoptic color mixing is obtained have already been outlined above. It therefore suffices to say that the extent of binocular summation for a given chromatic stimulus will depend on the task and the precise chromatic and spatiotemporal parameters of the stimulus.

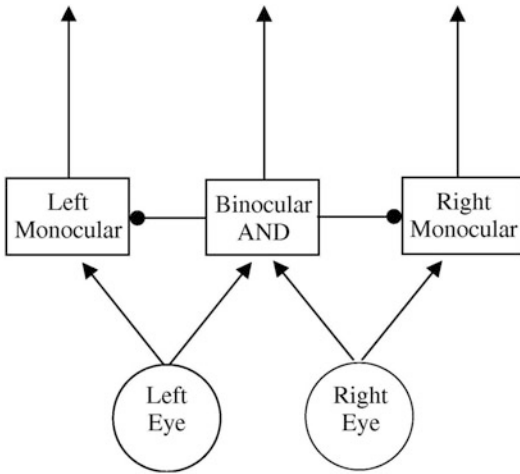
Binocular Color Appearance

Those of us who have typical binocular vision will have a region of binocular overlap in the middle of the visual field flanked on either side by regions that are only visible in one or the other eye. However, we are usually not aware that colors change in the center versus the periphery of our visual field, although changes are detectable under controlled laboratory conditions [3, 26]. A whole range of dichoptic mechanisms have been proposed to account for these apparent compensations between the two eyes. Shimono et al. [30] suggested a model which takes into account previous results on interocular transfer of chromatic adaptation and chromatic induction effects (e.g., [4, 8, 40]) by proposing a “purely binocular” chromatic mechanism, which responds only to simultaneous binocular presentation of similar colors to each eye but which also inhibits the responses of “purely monocular” color mechanisms (see Fig. 5). This model would also seem to be reasonable in the light of physiological data collected from monkey visual cortex [27]. However, more recent results, such as those of Kingdom and Libenson [14] and Mullen et al. [22] (discussed above), complicate this simple picture.

Another aspect of binocular color appearance that has recently received attention is how presenting different colors to the two eyes can be an easy way to simulate surface glossiness in stereoscopic visual displays, without the need to create a detailed reflectance model [13].



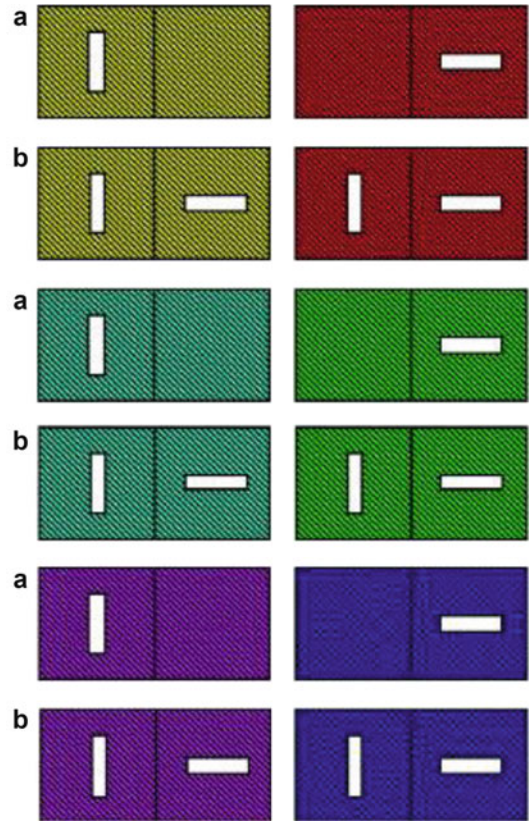
Binocular Color Matching, Fig. 4 Stimulus for chromostereopsis. If viewed on the full screen (especially if presented on a data projector), most people will see the red “CHROMO” look slightly in front of the blue “STEREOPSIS”



Binocular Color Matching, Fig. 5 Model from Shimono et al. [30] which illustrates the inhibitory action of their proposed “purely chromatic” binocular mechanism on monocular mechanisms (Taken from Shimono et al. [30], Reprinted from Vision Research, 49, Shimono, K., Shioiri, S., Yaguchi, H., Psychophysical evidence for a purely binocular color system, 202–210, Copyright (2009), with permission from Elsevier)

Conclusion

The field of binocular color vision has proven surprisingly controversial over the years, with the existence of true dichoptic color mixing, the contribution of color to stereopsis, and the presence or not of a purely binocular color mechanism being key battlegrounds. The current consensus of the field appears to be that dichoptic color mixing can happen, that chromatic mechanisms contribute to stereopsis, and that there is a purely binocular color mechanism, but demonstrating these things requires careful stimulus control and experimental technique. Added to this there are some curious phenomena such as being able to make stimuli which are easily detectable when viewed monocularly disappear when viewed binocularly [1] and the interocular transfer of chromatic induction [8] (Fig. 6). It seems that we still have some more work to do before we can say that we fully understand the complex interactions between binocular color mechanisms.



Binocular Color Matching, Fig. 6 Binocular color induction in monocular patches. Sets of multicolored stereograms are shown for three different color combinations. (a) The *white*, monocular patches gradually fill with distinct pastel colors during binocular viewing of the stereograms. The buildup of the effect typically takes a few seconds. Dependent on the viewer, the full development of the colors may even take minutes. The color effects are most compelling when the stereograms are viewed printed out and in bright light (see original publication). When one eye is closed, the illusion gradually becomes weaker and disappears. This buildup and decay of the illusion after opening or closing one eye indicates that the color appearances are induced by a binocular mechanism. (b) During binocular viewing, the *white*, binocular patches do not change color. During monocular viewing after a period of binocular viewing, however, induced colors are apparent in the patches (Taken from Erkelens and Van Ee [8], Reprinted from Vision Research, 42, Erkelens, C.J., Van Ee, R., Multi-coloured stereograms unveil two binocular colour mechanisms in human vision, 1103–1112, Copyright (2002), with permission from Elsevier)

Cross-References

- [Chromostereopsis](#)
- [Color Processing, Cortical](#)
- [Cone Fundamentals](#)
- [Ganglion Cells](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Hering, Karl Ewald Konstantin](#)
- [Magno-, Parvo-, and Koniocellular Pathways](#)
- [Simultaneous Color Contrast](#)
- [Stereo and Anaglyph Images](#)

References

1. Anstis, S., Rogers, B.: Binocular fusion of luminance, color, motion and flicker – two eyes are worse than one. *Vis. Res.* **53**, 47–53 (2012)
2. Bhardwaj, R., O'Shea, R.P.: Temporal analysis of image-rivalry suppression. *PLoS One.* **7**(9), e45407 (2012). <https://doi.org/10.1371/journal.pone.0045407>
3. Bompas, A., Powell, G., Sumner, P.: Systematic biases in adult color perception persist despite lifelong information sufficient to calibrate them. *J. Vis.* **13**, 1–19 (2013)
4. Coltheart, M.: Color-specificity and monocularity in the visual cortex. *Vis. Res.* **3**, 2595–2598 (1973)
5. De Weert, C.M.M., Wade, N.J.: Compound binocular rivalry. *Vis. Res.* **28**, 1031–1040 (1988)
6. Den Ouden, H.E.M., Van Ee, R.R., de Haan, E.H.E.: Colour helps to solve the binocular matching problem. *J. Physiol.* **567**, 665–671 (2005)
7. Desagulier, J.T.: A plain and easy experiment to confirm Sir Isaac Newton's doctrine of the different refrangibility of the rays of light. *Philos. Trans. R. Soc. Lond. A.* **29**, 448–452 (1716)
8. Erkelens, C.J., Van Ee, R.: Multi-coloured stereograms unveil two binocular colour mechanisms in human vision. *Vis. Res.* **42**, 1103–1112 (2002)
9. Grinberg, D.L., Williams, D.R.: Stereopsis with chromatic signals from the blue-sensitive mechanism. *Vis. Res.* **25**, 531–537 (1985)
10. Hovis, J.K.: Review of dichoptic color mixing. *Optom. Vis. Sci.* **66**, 181–190 (1989)
11. Howard, I.P., Rogers, B.J.: *Perceiving in Depth: volume 2 Stereoscopic Vision*. Oxford University Press, Oxford, UK (2012)
12. Jordan, J.R., Geisler, W.S., Bovik, A.C.: Color as a source of information in the stereo correspondence process. *Vis. Res.* **30**, 1955–1970 (1990)
13. Jung, W.-S., Moon, Y.-G., Park, J.-H., Song, J.-K.: Glossiness representation using binocular color difference. *Opt. Lett.* **38**, 2584–2587 (2013)
14. Kingdom, F.A.A., Libenson, L.: Dichoptic color saturation mixture: binocular luminance contrast promotes perceptual averaging. *J. Vis.* **15**(5), Article no. 2 (2015)
15. Kingdom, F.A.A., Simmons, D.R.: Stereoacuity and colour contrast. *Vis. Res.* **36**, 1311–1319 (1996)
16. Kingdom, F.A.A., Simmons, D.R.: The relationship between colour vision and stereoscopic depth perception (invited paper). *J. Soc. 3D Broadcast. Imaging.* **1**, 10–19 (2000)
17. Kingdom, F.A.A., Simmons, D.R., Rainville, S.R.: On the apparent collapse of stereopsis in random-dot stereograms at isoluminance. *Vis. Res.* **39**, 2127–2141 (1999)
18. Kovacs, I., Julesz, B.: Depth, motion, and static-flow perception at metaisoluminant color contrast. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10390–10394 (1992)
19. Livingstone, M.S., Hubel, D.H.: Psychophysical evidence for separate channels for the perception of form, colour, movement, and depth. *J. Neurosci.* **7**, 3416–3468 (1987)
20. Lu, C., Fender, D.H.: The interaction of color and luminance in stereoscopic vision. *Investig. Ophthalmol.* **11**, 482–489 (1972)
21. Malkoc, G., Kingdom, F.A.A.: Dichoptic difference thresholds for chromatic stimuli. *Vis. Res.* **62**, 75–83 (2012)
22. Mullen, K., Kim, Y.J., Gheiratmond, M.: Contrast normalization in colour vision: the effect of luminance contrast on colour contrast detection. *Sci. Rep.* **4**, 7350 (2014). <https://doi.org/10.1038/srep07350>
23. O'Shea, R.P., Williams, D.R.: Binocular rivalry with isoluminant stimuli visible only via short-wavelength-sensitive cones. *Vis. Res.* **36**, 1561–1571 (1996)
24. Ooi, T.L., Loop, M.S.: Visual suppression and its effect upon color and luminance sensitivity. *Vis. Res.* **34**, 2997–3003 (1994)
25. Ozolins, M., Muizniece, K.: Color difference threshold of chromostereopsis induced by flat display emission. *Front. Psychol.* **6**, Article: 337 (2015)
26. Parry, N.R.A., Panorgias, A., McKeefry, D.J., Murray, I.J.: Real-world stimuli show perceived hue shifts in the peripheral visual field. *J. Opt. Soc. Am. A.* **29**, 96–101 (2012)
27. Peirce, J.W., Solomon, S.G., Forte, J.D., Lennie, P.: Cortical representation of color is binocular. *J. Vis.* **8**, 1–10 (2008)
28. Sagawa, K.: Minimum light intensity required for color rivalry. *Vis. Res.* **21**, 1467–1474 (1981)
29. Scharff, L.V., Geisler, W.S.: Stereopsis at isoluminance in the absence of chromatic aberrations. *J. Opt. Soc. Am. A.* **9**, 868–876 (1992)
30. Shiono, K., Shioiri, S., Yaguchi, H.: Psychophysical evidence for a purely binocular color system. *Vis. Res.* **49**, 202–210 (2009)
31. Simmons, D.R.: The minimum contrast requirements for stereopsis. *Perception.* **27**, 1333–1343 (1998)
32. Simmons, D.R.: The binocular combination of chromatic contrast. *Perception.* **34**, 1035–1042 (2005)
33. Simmons, D.R., Kingdom, F.A.A.: Contrast thresholds for stereoscopic depth identification with

- isoluminant and isochromatic stimuli. *Vis. Res.* **34**, 2971–2982 (1994)
34. Simmons, D.R., Kingdom, F.A.A.: Differences between stereopsis with isoluminant and isochromatic stimuli. *J. Opt. Soc. Am. A.* **12**, 2094–2104 (1995)
 35. Simmons, D.R., Kingdom, F.A.A.: On the independence of chromatic and achromatic stereopsis mechanisms. *Vis. Res.* **37**, 1271–1280 (1997)
 36. Simmons, D.R., Kingdom, F.A.A.: Interactions between chromatic- and luminance-contrast-sensitive stereopsis mechanisms. *Vis. Res.* **42**, 1525–1546 (2002)
 37. Simmons, D.R.: Stereopsis at red-green isoluminance: chasing the luminance artifacts. [Abstract]. *J. Vis.* **2**(10), 69a (2002). <https://doi.org/10.1167/2.10.69>. <http://journalofvision.org/2/10/69/>
 38. Stuart, G.W., Edwards, M., Cook, M.L.: Color inputs to random-dot stereopsis. *Perception.* **21**, 717–729 (1992)
 39. Tyler, C.W., Cavanagh, P.: Purely chromatic perception of motion in depth: two eyes as sensitive as one. *Percept. Psychophys.* **49**, 53–61 (1991)
 40. Vimal, R.L.P., Shevell, S.K.: A central binocular mechanism affects chromatic adaptation. *Vis. Res.* **27**, 429–439 (1987)
 41. Wardle, S.G., Gillam, B.J.: Color constrains depth in da Vinci stereopsis for camouflage but not occlusion. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1525–1540 (2013)
 42. Yoonessi, A., Kingdom, F.A.A.: Dichoptic difference thresholds for uniform color changes applied to natural scenes. *J. Vis.* **9**(2), 1–12 (2009) (article no. 3)
 43. Krauskopf, J., Forte, J.D.: Influence of chromaticity on vernier and stereo acuity. *J. Vis.* **2**(9), 645–652 (2002)

Binocular Color Perception

► Binocular Color Matching

Binocular Color Vision

► Binocular Color Matching

Biopolymers

► Inorganic, Hybrid, and Functional Pigments

Black Body

► Blackbody and Blackbody Radiation

Blackbody and Blackbody Radiation

Joanne Zwinkels

National Research Council Canada, Ottawa, ON, Canada

Synonyms

Blackbody (specific terms) – [Black body](#); [Full radiator](#); [Natural object](#); [Planckian radiator](#)

Blackbody (alt. terms) – [Ideal black surface](#); [Ideal emitter](#); [Ideal thermal radiator](#)

Blackbody radiation – [Black-body radiation](#); [Cavity radiation](#)

Blackbody locus – [Planckian locus](#)

Definition

A *blackbody* is an ideal thermal radiator that absorbs completely all incident radiation whatever the wavelength, the direction of incidence, or the polarization [1]. This radiator has, for any wavelength and any direction, the maximum spectral concentration of radiance for a thermal radiator in thermal equilibrium at a given temperature.

Blackbody radiation is radiant energy emitted by an ideal black surface (blackbody) whose spectral power distribution is only governed by its own temperature.

Blackbody color is the temperature of a true blackbody emitter which to the human eye is a close match to the color of an incandescent object, which is not a blackbody.

Blackbody locus is the locus of points in a chromaticity diagram that represents chromaticities of the radiation of Planckian radiators at different temperatures.

Description

All bodies at nonzero temperature absorb and emit radiation. If a body is at thermal equilibrium, it is because the amount of radiation that is emitted per second is equal to the amount absorbed. A blackbody (or ideal Planckian or black radiator) is a body that can completely absorb radiation at any wavelength and completely emit radiation at any wavelength so that it emits more radiant power at each wavelength than any other object at the same temperature. A blackbody is not only an ideal emitter but also a diffuse emitter, which means that the emitted radiation is isotropic or independent of direction [2]. It gets its name because of the fact that, at room temperature, it appears visibly black.

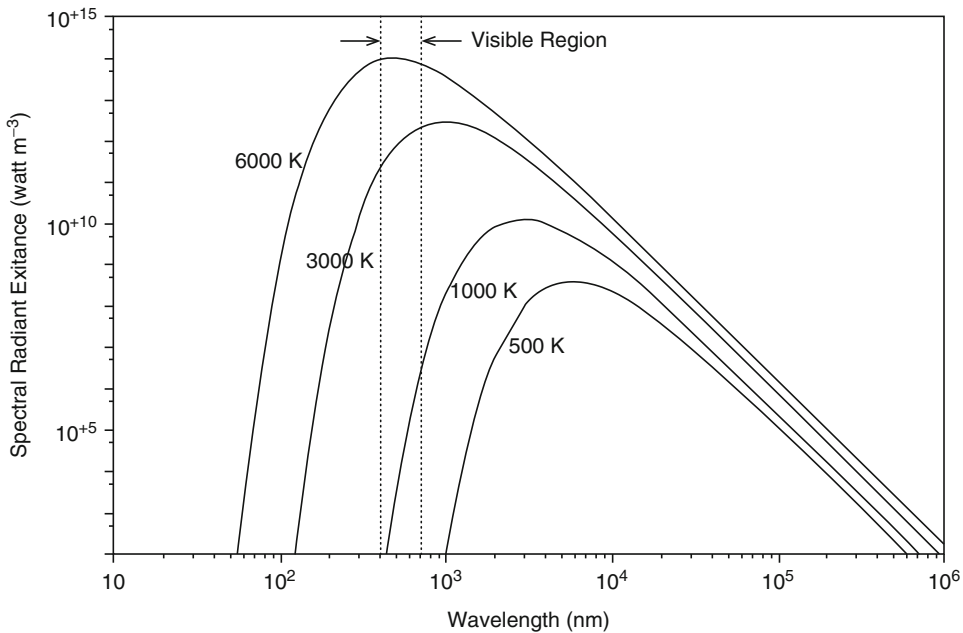
In the late nineteenth century, Kirchhoff [3] and other experimental physicists found that an ideal blackbody, in thermodynamic equilibrium, emitted radiant energy whose spectrum was continuous and a function only of its wavelength and temperature, independent of the nature of the blackbody. However, this behavior could not be explained using thermodynamics or classical

statistical mechanics. It was not until Einstein introduced the idea that light could be quantized that Planck formulated his blackbody radiation law based upon this principle [4]. This is generally known as *Planck's law* and gives the amount of power emitted by a blackbody in equilibrium at a temperature T according to:

$$M_e(\lambda, T) = \frac{c_1}{\lambda^5} \left(\frac{1}{e^{c_2/\lambda T} - 1} \right)$$

where $M_e(\lambda, T)$ is the spectral radiant exitance (power per unit area per unit wavelength interval), λ is the wavelength in meters, and T is the absolute temperature in kelvins. The constants c_1 and c_2 are known as Planck's radiation constants and have the values: $c_1 = 2\pi hc^2 = 3.74177 \times 10^{-16} \text{ W m}^2$; $c_2 = hc/k = 1.43878 \times 10^{-2} \text{ m K}$, where h is Planck's constant, c is the speed of light in vacuum, and k is Boltzmann's constant.

The spectral power distributions of some blackbody radiators at different temperatures are shown in Fig. 1. It can be seen that the peak of the blackbody radiation curve shifts to shorter wavelength and higher energy with increasing



Blackbody and Blackbody Radiation, Fig. 1 Blackbody spectral exitance curves for several temperatures

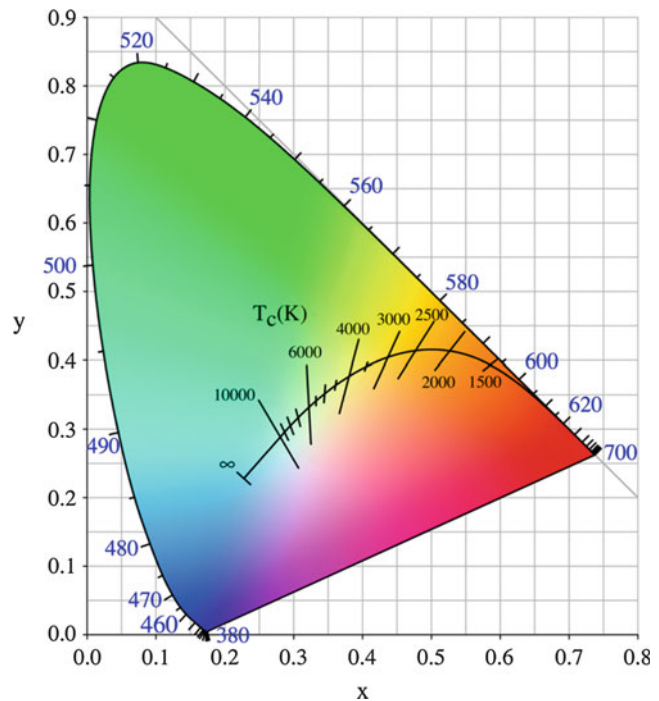
temperature. At room temperature (300 K), the peak of the curve is at about 10 μm , and the emission is largely thermal in the infrared region. As discussed above, this room-temperature blackbody that has essentially no emission in the visible region appears black. However, as the temperature of the blackbody source rises through the range shown, it increasingly emits in the visible region, and its color changes continuously from black to red through orange and yellow to white and finally a bluish white.

The color coordinates of these blackbody sources at different temperatures can be plotted on a chromaticity diagram. This is shown in Fig. 2, which shows that as the temperature of the blackbody radiator increases, its CIE chromaticity coordinates (x, y) move along a curved line in the CIE 1931 chromaticity diagram. This curve is known as the *blackbody* or *Planckian locus*. For white light sources for general lighting applications, they are designed so that their CIE chromaticities (u^*v^*) are located near this Planckian locus, when plotted on a Uniform Color Scale (UCS) diagram (u^*v^*).

For colorimetric applications, the International Commission on Illumination (CIE) has defined a standard illuminant A, which is representative of illumination under incandescent lighting [5] and whose values of relative spectral power distribution are that of a Planckian radiator at a temperature of approximately 2856 K. It is important to note that while the original CIE standard illuminant A was defined in 1931 in terms of a relative spectral power distribution of a Planckian radiator at a temperature of 2848 K, the Planckian radiation constants, c_1 and c_2 , have changed several times since then (in 1948, 1968, and 1990) with changes in the International Temperature Scale (ITS), causing a corresponding change in the associated temperature. For this reason, the CIE Standard which currently defines CIE standard illuminant A is no longer a function of temperature [5]. However, for an artificial source that is representative of CIE standard illuminant A, the concept of *correlated color temperature* is used, which is the temperature of the Planckian radiator having a chromaticity along the Planckian locus nearest the chromaticity calculated for the spectral

Blackbody and Blackbody Radiation,

Fig. 2 CIE 1931 chromaticity diagram showing the locus of blackbody radiators calculated for Planckian temperatures from 1500 K to infinity; the Planckian temperature that provides the smallest chromaticity difference between an artificial source and this blackbody locus is its correlated color temperature, T_c



distribution of the test source. There are also many artificial white light sources whose spectral power distribution is very different from a blackbody source, such as fluorescent lamps and sources based on light-emitting diodes. However, the concept of correlated color temperature is still used to describe the colorimetric properties of these types of white light sources.

For astronomical applications, the electromagnetic radiation from stars and planets is sometimes characterized in terms of an *effective temperature*, the temperature of a blackbody that would emit the same total flux of electromagnetic energy. However, stars are observed at a depth inside the atmosphere (the photosphere), so it is more typical to describe a star's visual surface by its *photospheric temperature* (see entry on ► [“Apparent Magnitude, Astronomy”](#)).

There are several other important properties of blackbody sources that are given by the following three laws [6]:

Stefan-Boltzmann Law: This law states that the total radiant exitance, M of a blackbody, determined by integrating the spectral radiant exitance curve, $M_e(\lambda)$, for all wavelengths, λ , is proportional to the fourth power of its absolute temperature:

$$M = \sigma T^4$$

where σ is Stefan's constant ($=0.567037 \times 10^{-7} \text{ Wm}^2 \text{ K}^{-4}$). No surface at thermal equilibrium can emit more than this amount of energy.

Kirchhoff's Law of Thermal Radiation: This law states that, at each wavelength, the emissivity of the blackbody, $\varepsilon(\lambda)$, at thermal equilibrium is equal to its absorptivity, $\alpha(\lambda)$:

$$\varepsilon(\lambda) = \alpha(\lambda)$$

Wien's Displacement Law: This law states that the curves of the spectral power distribution of the blackbody radiation are peaked at a wavelength, λ_{peak} , given by:

$$\lambda_{\text{peak}} = \frac{2898}{T}$$

This can be simply shown by differentiating the Planckian spectral power distribution with respect to wavelength and solving the derivative when it equals zero.

Examples

Since the last century, blackbody radiation has been of great interest for use as a primary (calculable) source because the properties are universal and independent of the particular material substance. A good laboratory implementation of a blackbody is a small hole in the wall of an otherwise opaque cavity [7]. The light that enters the cavity through the hole will undergo so many reflections before it can escape from the hole that the light will be effectively absorbed and the hole will appear black. Until 1979, a blackbody like this, heated to the temperature of melting platinum, was the basis of the definition and realization of the SI base unit for photometry, the candela [8].

The earliest types of artificial light were based on heating an object until it was red hot or hotter and was able to emit visible light. These types of artificial light sources are referred to as *incandescent*. Examples of incandescent objects are fire, candles, carbon arcs, and tungsten-filament lamps. They emit radiation with approximately the same spectral power distribution as a blackbody but with an intensity reduced by a factor called the *emissivity*, ε . If the emissivity does not vary with wavelength, the source is called a *gray body* or *nonselective radiator*. An incandescent tungsten-filament lamp is close to being a gray body with an emissivity of approximately 0.40 [7]. The temperature of a blackbody that is closest in color to this gray body source is often used as an index of its color. This is referred to as the *correlated color temperature*, T_c . An alternative index is the *distribution temperature*, which is the temperature of a blackbody whose relative spectral power distribution is closest to that of the source in question. However, all three terms should be

used with caution as they do not give a full picture of the actual spectral distribution [9]. Typical tungsten-filament lamps have correlated color temperatures and distribution temperatures in the range of 2500–3000 K.

However, these incandescent lamps emit not only visible light but a significant amount of infrared radiation which makes them inefficient as light sources compared with more modern lighting technologies, such as solid-state lighting which can be designed to emit optical radiation only in the visible range.

There are many examples of everyday materials that behave like blackbodies but emit significant thermal radiation. The human body is one example of a blackbody radiator which emits at roughly 300 K. Another obvious example is the Sun, whose surface is roughly 6000 K. The Earth is also considered to be a blackbody radiator, but it is less than ideal.

Cross-References

- [Apparent Magnitude, Astronomy](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [Spectral Power Distribution](#)
- [USC Diagrams; Uniform Chromaticity Scales; \$Y_u/v'\$](#)

References

1. CIE S017/E:2020, ILV – International Lighting Vocabulary, 2nd ed., CIE Central Bureau, Vienna
2. Mahmoud, M.: §2.1 Blackbody radiation. In: *Engineering Thermofluids: Thermodynamics, Fluid Mechanics, and Heat Transfer*, p. 568. Springer, Berlin (2005). ISBN 3–540–22292-8
3. Kirchhoff, G.U.: "Über das Verhältniss zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme und Licht. *Poggendorff Annalen der Physik und Chemie*, 1860, v. 109, 275–301. (English translation by F. Guthrie: Kirchhoff G. On the relation between the radiating and the absorbing powers of different bodies for light and heat. *Phil. Mag.*, ser. 4, v. 20, 1–21 (1860))
4. Planck, M.: Distribution of energy in the spectrum. *Ann. Phys.* **4**, 553–563 (1901)
5. ISO 11664-2:2007/CIE, S014-2:2006 Colorimetry – Part 2: CIE standard illuminants for colorimetry
6. Williamson, S.J., Cummins, H.Z.: *Light and Color in Nature and Art*. Wiley, New York (1983)
7. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
8. Zwinkels, J.C., Ikonen, E., Fox, N.P., Ulm, G., Rastello, M.L.: Photometry, radiometry and "the candela" evolution in the classical and quantum world. *Metrologia*. **47**, R15–R32 (2010)
9. Robertson, A.R.: Computation of correlated color temperature and distribution temperature. *J. Opt. Soc. Am.* **58**, 1528–1535 (1968)

Black-body Radiation

- [Blackbody and Blackbody Radiation](#)

Blacklight Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Ultraviolet radiator](#); [Wood's lamps](#)

Definition

Radiators that emit UV-A radiation together with very little visible radiation.

Ultraviolet Radiators and Blacklight Lamps

Many different ultraviolet radiators, based on both gas discharge lamp and light-emitting diode technologies, are produced for industrial applications such as photochemical processes, drying and hardening of materials, lithography, and reprography; for disinfection purposes and insect repellents; and for phototherapy and sun tanning. These radiators are available in types that radiate in the UV-A (wavelength range, 315–400 nm),

Blacklight Lamp,

Fig. 1 Blacklight show: different fluorescent dyes used in the clothing material light up in different colors under UV-A blacklight lamps (Photograph: Blacklight Theatre Prague: Creative Commons Attribution-ShareLike 4.0 International)



UV-B (280–315 nm), and UV-C (100–280 nm) part of the spectrum. In the applications listed above, visual effects play no role. Another type of ultraviolet radiators is produced to make fluorescent substances in materials visible to the eye. These radiators produce for this purpose UV-A radiation combined with a very small part of visible radiation, which are mainly dealt with here. They are usually called blacklight lamps. Blacklight lamps are used for artistic and decorative reasons (especially in theaters, discos, and bars; Fig. 1) and for making fluorescent dyes visible in banknotes, stamps, and documents to detect counterfeits, for forensic crime scene inspections, and for making fluorescent bacterial infections visible for medical purposes. Some artists have specialized themselves in using fluorescent paints. Their paintings are meant to be seen under blacklight lamps (Fig. 2).

Working Principle and Properties of Blacklight Lamps

Three different types of light sources are used to produce blacklight lamps:

- Low-pressure mercury lamps (both in tubular and compact shape).

- High-pressure mercury lamps.
- Light-emitting diodes (LEDs).

For the mercury-based blacklight radiators, the only difference compared with the mercury lamp executions that produce mainly visible light is the type of glass and the type of fluorescent powder used. As discussed in the relevant lamp chapters (low-pressure mercury lamps and high-pressure mercury lamps), these lamps produce ultraviolet radiation as a result of the discharge between two electrodes. The glass of the discharge tube of the blacklight versions is produced either of Wood's glass that transmits ultraviolet and infrared radiation but blocks most visible radiation (with the exception of violet around 400 nm and deep red around 800 nm) or of regular glass coated with a filter that transmits ultraviolet and blocks most of the visible radiation. Wood's lamps were introduced in the First World War as invisible signal beams used for communication [1]. The coated regular glass lamps emit UV-A radiation together with violet-bluish visible light. Their spectrum depends on the type of fluorescent powder used on the interior of the discharge tube. Either barium disilicate (industrial designation "BSP") with a resulting peak wavelength at 349 nm or strontium tetraborate (industrial designation "SBE") with a peak wavelength at 368 nm is used. The high-

Blacklight Lamp,

Fig. 2 Paintings by the artist Beo Beyond produced with fluorescent paint displayed in the artist's own blacklight studio in Barcelona, Spain (Photograph: Beo Beyond: Creative Commons Attribution-Share-Alike 3.0 Unported)



pressure mercury type of blacklight lamp exists in fluorescent-coated (SBE) and clear versions. The fluorescent version has a higher radiant flux in UV-A, while the clear version has a more concentrated light-emitting area making it more suitable for use in reflector type of luminaires. The high-pressure mercury blacklight lamps are more compact than the low-pressure mercury lamps when comparing the same ultraviolet radiation output. They can, therefore, be used in floodlight type of luminaires as used for theatrical and concert performances. They are also available in higher output versions. The low-pressure mercury versions are available in the wattage range from 4 to 36 W with UV-A radiation output power between some 1 and 10 W. The high-pressure versions are available in the wattage range from 100 to 1000 W with UV-A radiation output between some 50 and 500 W.

Blacklight LEDs that emit radiation around the 390–420 nm wavelength range are increasingly replacing the mercury-based blacklight lamps. The semiconductor material used for UV-A LEDs is gallium nitride (GaN) or aluminum gallium nitride (AlGaN). UV-LEDs emit the UV radiation directly without the use of phosphors

or filters. They have a small spectrum (quasi-monochromatic). Blacklight LEDs are available in a wide range of wattages, lumen packages, and intensities. They are produced in spotlight as well as strip or string versions and in many different dimensions. The string versions are used to combine the fluorescent effect on radiated materials in a space with the decorating effect of the purple glowing string, which can be shaped in all kind of forms.

Safety

Blacklight lamps radiate part of their light in the upper range of the UV-A spectrum which borders onto the visible part of the spectrum. The UV-A part of the total range of UV radiation (UV-C, UV-B, and UV-A) is the least hazardous part of the ultraviolet spectrum. Hazardous here relates to the risk of damage of the human skin and product materials. Radiation with wavelengths nearest to the visible part of the spectrum is, in turn, the least hazardous part of the total UV-A range. It is the combination of radiation power installed, the wavelengths that the actual blacklights radiate,

and the time of radiation exposure that determines possible hazards. For situations where the blacklight source is in the direct line of sight, also the radiance (the radiation equivalent for luminance) plays an essential role. Many blacklight installations are safe, if care is taken that the blacklight sources are outside the direct lines of sight. But, the first three aspects listed above vary from installation to installation so that it is wise to check if safety limits are not exceeded. Especially evaluating the exposure time for the working staff, who may be exposed for periods of around 8 hours, may be necessary in this context.

Cross-References

- [High Pressure Mercury Lamp](#)
- [Light-Emitting Diode, LED](#)
- [Phosphors and Fluorescent Powders](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Dieke, G.H.: Robert Williams Wood. Biographical Memoir, National Academy of Sciences, Washington DC (1993)

Blending

- [Compositing and Chroma Keying](#)

Blind Spot

Jonathan Aboshiha
The UCL Institute of Ophthalmology and
Moorfields Eye Hospital, London, UK

Synonyms

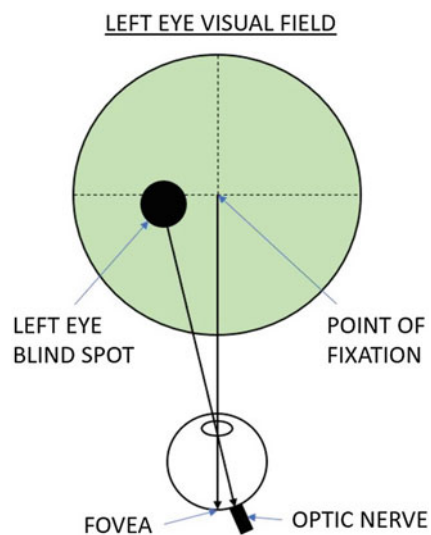
[Physiologic scotoma](#); [Punctum cecum](#)

Definition

Blind spot is the name given to the scotomatous area of each eye's visual field that lacks visual input, due to the photoreceptor-free region of the retina where the optic nerve exits the eye.

Overview

The phenomenon of the blind spot was first described in 1668 by the French Roman Catholic priest and scientist Edme Mariotte, who used small circles of white paper to locate the region of visual space in which they disappeared [1]. Prior to this discovery, the role of the optic disc was not clear, and scientists such as Leonardo da Vinci (1452–1519) had thought that the visual image actually fell only on the head of the optic nerve itself [2]. The vertically oval area of the blind spot subtends about $5 \times 7^\circ$ of visual angle and corresponds to the projection in visual space of the optic nerve head, which has average vertical and horizontal disc diameters of 1.88 and 1.77 mm, respectively [3]. The physiological blind spot is centered about 15° temporally from the point of fixation, due to the nasal anatomical location of the optic nerve head in relation to the



Blind Spot, Fig. 1

fovea (Fig. 1). One is normally not aware of the area of the visual field which corresponds to the blind spot in binocular vision, due to the fellow eye's retinal and cortical representation of the corresponding area of the visual field. Even in monocular vision, this region is perceptually filled in at a cortical level using the brightness, color, and texture detected in the visual field adjacent to the blind spot [4]. This filling in of the scotomatous area of visual field caused by the optic nerve head has also been shown to be an active process involving spatial integration, wherein the pattern and color of not just the area immediately adjacent to the scotoma but also from the more remote surround are used to perceptually fill in the scotomatous region [5].

The size of the blind spot strongly correlates with the area of the optic disc, as well as other features such as the region around the nerve head where the retina has wasted away (i.e., parapapillary chorioretinal atrophy) [6] and the surface topography of the disc, with the more prominent nasal part of the disc appearing less "blind" than the shallow temporal part, probably due to a higher degree of light scattering by the more prominent nasal disc region [7]. In certain pathological conditions, most of which affect the optic nerve head, such as inflammation (i.e., optic neuritis) or optic disc swelling due to raised intracranial pressure, the area of the blind spot can become enlarged, and this can be detected clinically using automated visual field tests or, in more gross abnormalities, confrontational estimation of the blind spot size using a small target such as red hatpin in direct comparison with a normal observer's blind spot. Such pathological conditions of the optic nerve are often associated with defects in color vision, most classically that of red desaturation. These defects may be further evident on testing using pseudo isochromatic plates, typically demonstrating more severe red-green deficits with milder blue-yellow losses [8]. Only the vertebrate eye has a physiological blind spot, on account of light having to pass through the anterior retinal nerve fiber layer (which forms the optic nerve head as it exits the

eye) in order to reach the photoreceptor layer beneath; cephalopods, such as the squid and octopus, have retinæ wherein the nerve fiber layer lies behind the photoreceptor layer and so have no physiological blind spot [9].

Cross-References

► [Pseudoisochromatic Plates](#)

References

1. Grzybowski, A., Aydin, P.: Edme Mariotte (1620–1684): pioneer of neurophysiology. *Surv. Ophthalmol.* **52**(4), 443–451 (2007)
2. Arrington, G.E.: *A History of Ophthalmology*. MD Publications, New York (1959)
3. Quigley, H.A., Brown, A.E., Morrison, J.D., Drance, S. M.: The size and shape of the optic disc in normal human eyes. *Arch. Ophthalmol.* **108**(1), 51–57 (1990)
4. Spillmann, L., Otte, T., Hamburger, K., Magnussen, S.: Perceptual filling-in from the edge of the blind spot. *Vis. Res.* **46**(25), 4252–4257 (2006)
5. Ramachandran, V.S., Gregory, R.L.: Perceptual filling in of artificially induced scotomas in human vision. *Nature*. **350**(6320), 699–702 (1991)
6. Jonas, J., Gusek, G., Fernández, M.: Correlation of the blind spot size to the area of the optic disk and parapapillary atrophy. *Am J. Ophthalmol.* **111**(5), 559–565 (1991)
7. Meyer, J.H., Guhlmann, M., Funk, J.: Blind spot size depends on the optic disc topography: a study using SLO controlled scotometry and the Heidelberg retina tomograph. *Br. J. Ophthalmol.* **81**(5), 355–359 (1997)
8. Yates, J., Diamantopoulos, I., Daumann, F.: Acquired (transient and permanent) color vision disorders. In: Menu, J. (ed.) *Operational Color Vision in the Modern Aviation Environment*, p. 1. North Atlantic Treaty Organization Research and Technology Organization A, Neuilly-Sur-Seine Cedex (2001)
9. Miller, R.F.: The physiology and morphology of the vertebrate retina. In: Ryan, S.J. (ed.) *Retina*, vol. 1, 2nd edn, pp. 58–71. Mosby, St. Louis (1994)

"Blue-Blindness"

► [Tritanopia](#)

Blue-Yellow Deficiency

- ▶ [Tritanopia](#)

Brightness Perception Associated with Photoreceptors' Stimulations

- ▶ [Melanopsin Contributions to Human Brightness Perception](#)

Brightness Perception Based on Physiology

- ▶ [Melanopsin Contributions to Human Brightness Perception](#)

Brightness-Luminance Ratio: B/L

- ▶ [Helmholtz-Kohlrausch Effect](#)

Canonical Color

► [Memory Color](#)

Carbon Arc Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Arc lamps](#)

Definition

Lamps consisting of two rods of carbon in the open air or in a glass enclosure. The ends of the rods touch each other and are connected to a current source. By subsequently separating the rods, a discharge arc is produced that brings the ends of the rods to bright incandescence.

Carbon Arc Electric Lamps

In contrast to what many people think, it is not the incandescent lamp, but the carbon arc lamp that was the first electric light source used.

Already in 1810, Humphry Davy demonstrated in the Royal Institution in London a bright arc between two pieces of charcoal connected to 2,000 voltaic cells [1–3]. Electric carbon arc lighting really took off after the introduction of steam-driven generators around 1850, some 30 years before the introduction of the incandescent lamp. The earliest practical application of electric light was an arc lamp used to simulate the sun in the opera of Paris in 1849 [1]. Arc lamps with their concentrated light of high intensity were, in their infancy, especially used for beacon and searchlights. From 1870 onwards, arc lamps became popular for the lighting of streets, factory halls, railway stations, and big department stores. Huge structures, sometimes called moonlight towers, were used in cities to illuminate large areas instead of using many small masts with gas lanterns. The arc lamp had such a high intensity that it was seldom used in domestic lighting.

From the beginning of the twentieth century, incandescent electric lighting quickly replaced carbon arc lighting installations. Up to the 1950s, extremely high-intensity arc lamps were still used in searchlights, in film studios, and in cinema projectors, until the short-arc gas discharge xenon lamp took over. Today, electrical arcs are used not for lighting but for industrial purposes, as, for example, in plasma torches and welding apparatus where an arc is created between the one welding rod and the metal material to be welded.

Working Principle

When two pointed carbon rods connected to an electric current source touch each other, the resistance at the pointed ends is so high that the rods are heated and begin to glow. When subsequently the rods are separated, they are warm enough for the negatively charged one to easily emit electrons: a discharge is created between the two rods. Usually, the carbon rods are referred to as electrodes; the negative charged one, the cathode, and the positively charged one, the anode. The electrons of the discharge move from the negative to the positive carbon electrode and bombard the anode, heating it. The largest part of the bright light does not come from the arc discharge itself but from the end of the electrodes which are brought to incandescence. The heated air around the discharge rises and makes the bright area rise in the form of an arch giving the lamp its name of arc lamp (Fig. 1).

The gap between the rods is just a few millimeters, and the light-emitting area, therefore, is so small that concentrated light of high intensity is created. The carbon rods burn away with time; in a DC supply, the positive rod burns more quickly than the negative rod because it becomes hotter. The distance between the rods has to be adapted regularly as the arc will extinguish if the distance

becomes too large. Many different mechanisms have been invented to perform this automatically. After some time, the rods become so short that they have to be replaced.

An arc lamp has a negative-resistance characteristic (like all gas discharge lamps) and needs, therefore, a resistor, usually an inductive coil, in its electric circuit to limit the current.

Materials and Construction

Electrodes

Common Carbon Rods

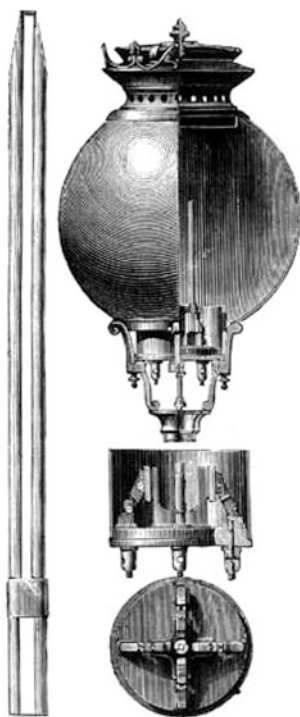
Charcoal was initially used for the electrodes, but charcoal burns away rapidly. It was soon discovered that rods made out of carbon have a much longer life. Hard molded carbon rods were therefore used which later got a core of soft carbon. DC-operated lamps used for the anode a thicker carbon rod than for the cathode to make them burning away with the same rate. In AC-operated lamps, the burning rate of some 20 mm per hour of anode and cathode is, of course, the same [1, 3]. The rods have a diameter of 10 mm to slightly more than 15 mm and were made as long as possible, up to some 500 mm, giving a lifetime of up to 24 h.

Jablochkoff's Parallel Electrodes

Around 1880, the Russ Paul Jablochkoff introduced a whole new concept of electrodes that did away with the need for continuous regulation of the distance between the rods. The "Jablochkoff electric candle," as it is usually called, consists of two parallel rods of carbon separated by plaster (Fig. 2). For ignition, a bridge piece of carbon is positioned at the top. The plaster functions as an electric isolator between the rods and restricts the arc to the top of the electrodes. The plaster crumbles off as the carbon burns down. The position of the light-emitting area moves down with the burning of the candle, making these devices unsuitable for projection type of applications. Since the candle was burned up in 1–2 h, automatic replacement mechanisms for the candles were introduced.



Carbon Arc Lamp, Fig. 1 Because of the rise of heated air, the arc rises in the form of an arch. (Photograph: Achgro: Creative Commons 3.0 unported)



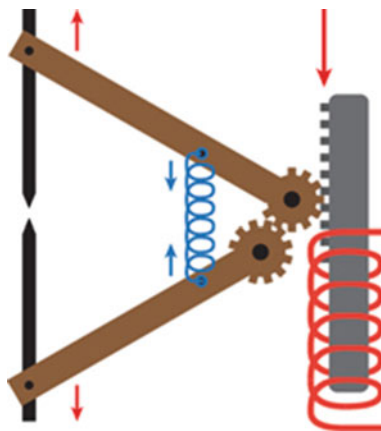
Carbon Arc Lamp, Fig. 2 Jablochkoff parallel carbon electrodes, separate and as used in an enclosed lantern (Drawing 1876) [6]

Carbon Rods with Additives (Flame Arc Lamps)

Just before 1900, fluorides of certain metals (including rare earth metals) were added to the carbon rods. When the electrodes become hot, the metallic salts evaporate and take part in the arc discharge, enveloping the arc as a flame, hence the name of flame arc. Both the lumen output and the luminous efficacy increase considerably with a factor between 2 and 4. These types are therefore also referred to as “high-intensity arc lamps.” Rare earth additives emit a line spectrum resulting in bright white light. Other types of additives emit different colors of light as, for example, calcium, emitting explicitly yellow light, and strontium, red light. In this way, light sources emitting specific spectra suitable for chemical and photographic processes were produced [3].

Electrodes Regulator

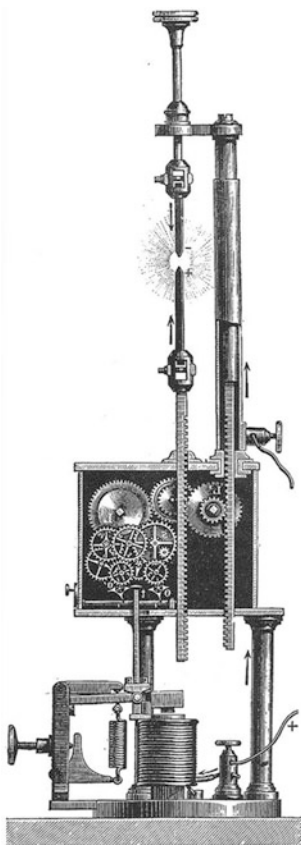
As has been mentioned, the distance between the rods has to be adapted regularly as the arc will



Carbon Arc Lamp, Fig. 3 The principle of a self-regulating mechanism making use of the force of a spring (blue) and that of an electromagnet (red)

extinguish if the distance becomes too large in the process of burning off material from the rods. Simple hand-regulated devices were designed where, by turning one screw, both electrodes were adjusted so that the light-emitting area remained at the same location. These systems have long been used in arc lamps for cinema projection.

Early self-regulating mechanisms made use of a clockwork winding device. Later mechanisms use the force of electromagnets. The force of an electromagnet, put in the same circuit as the rods, pushes the rods apart (Fig. 3). At the same moment, the rods are pulled together by gravity force or, as in Fig. 3, by spring force. When the gap between the rods increases, the resistance in the circuit increases and the current therefore decreases. Because of the decreased current, the pushing force of the electromagnet decreases as well, so that the gap size and gap position remain unchanged. The same mechanism takes care of automatic ignition when the power is turned on. When the power is switched off, the rods move to each other until they touch because of the spring force. When the power is switched on again, the large current through the system and thus through the electromagnet moves the rods from each other against the spring force, igniting the lamp automatically. For accurate control,



Carbon Arc Lamp, Fig. 4 A self-regulating arc lamp, after Foucault, balancing the force of gravity with the force of an electromagnet [1]

sometimes complicated clockwork types of gears were applied (Fig. 4).

Lantern

Enclosed Arc

Around 1900, the enclosed arc was introduced with which the lifetime of the carbon rods was increased with a factor of more than five. In a glass globe surrounding the arc, the oxygen is rapidly consumed by the burning electrodes and thereafter the carbon is burned away much slower. Burning times of up to 150 h are possible without rod replacement [1, 3]. Both the light output and the efficacy of the enclosed arc lamp are slightly

lower than that of the open arc lamp. Figure 5 shows a page of a catalogue with some enclosed carbon arc street-lighting lanterns from the early last century.

Optics

For low-mast street-lighting applications and for indoor industrial applications, lanterns usually employed opal or prismatic glass covers. The compact high-intensity light of arc lamps makes them preeminently suitable for floodlighting, for signal lights (in lighthouses, for example), and for searchlights. For this purpose, advanced mirror optical systems were designed (Fig. 6).

Properties

Open carbon arc lamps have a light output of up to some 4,000 lm (500 W versions) at a luminous efficacy of some 4–8 lm/W. Enclosed lamps have a 10–15% lower output and efficacy. Flame arc lamps, using carbon rods with additives, have a light output up to 15,000 lm (500 W versions) at efficacies between 15 and 30 lm/W [3]. Some arc lamps designed for use in searchlights have wattages of more than 20 kW. Beam intensities of up to 5,000 million candelas have been reported with mirror diameters of more than 2 m.

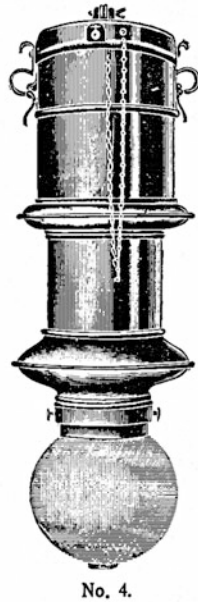
Arc lamps are often not rated by power but by the current they draw. Lamps with currents from 5 to 1,000 amp have been produced.

The correlated color temperature of some 3800 K of arc lamps [5] is much higher than what one was accustomed to with oil, candle, and gas lighting. The high-intensity flame arc lamps have relatively high color temperatures, depending on the material, up to 5,000 K. The spectrum of flame arc lamps extends well into the ultraviolet part (UV-A, B, and C) so that care is required with open arc lamps. Special arc lamp devices for tanning purposes have in fact been produced.

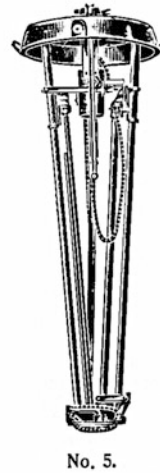
Arc lamps produce a buzzing sound which in interiors was experienced as annoying.

Carbon Arc Lamp,
Fig. 5 Carbon arc street-lighting lantern with self-regulating gap distance between the arcs [4]

Aussenansicht der
Type G. u. W. 60
mit runder Glocke



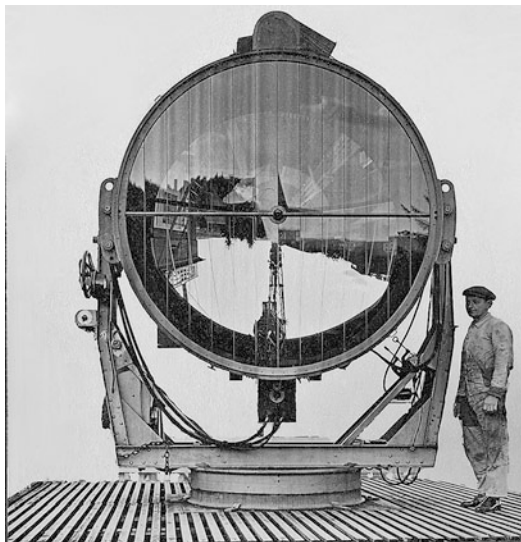
Innenansicht der
Lampentype G. u. W. 60



Aussenansicht der
Type G. u. W. 60
mit ovaler Glocke



Mit Erscheinen dieser Preisliste werden sämtliche früheren
Preislisten und Offerten ungültig.



Carbon Arc Lamp, Fig. 6 Carbon arc searchlight with a parabolic mirror of 2 m diameter, with chief mechanics Heinrich Beck and Erich Koch beside it (1930s). (Photograph: Heinrich Beck Institut, Germany)

Cross-References

► [Xenon Lamp](#)

References

1. Stoer, G.W.: History of Light and Lighting. Philips Lighting Division, Eindhoven (1986)
2. Rebske, E.: Lampen Laternen Leuchten. Franckh'sche Verlagshandlung, Stuttgart (1962)
3. Luckiesh, M.: Artificial Light Its Influence Upon Civilization. The Century Books of Useful Science, New York (1920)
4. Thirring, H.: Handbuch der Physik. Springer, Berlin (1928)
5. Macbeth, N.: Color temperature classification of natural and artificial illuminants. Trans. IES. **23**, 302–324 (1928)
6. Figuier, L.: Les merveilles de la science ou description des inventions modernes, Tome 2. Frurne, Jouvett et Cie (1891)

CAT

- [CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02](#)
-

Catalysts

- [Inorganic, Hybrid, and Functional Pigments](#)
-

Categorical Color Constancy

- [Color Constancy and Categories](#)
-

Categorical Perception

- [Effects of Color Terms on Color Perception and Cognition](#)
-

Cavity Radiation

- [Blackbody and Blackbody Radiation](#)
-

Chevreul, Michel-Eugène

Georges Roque
Centre National de la Recherche Scientifique
(CNRS), Paris, France

Definition

Michel-Eugène Chevreul (1786–1889) is one of the most important chemists of nineteenth-century France. A pioneer of organic chemistry, he was

twice President of the French Academy of Sciences. His work changed dramatically after his appointment as director of the dyeing department of the Gobelins Manufacture in Paris, where he worked for almost 60 years. At the Gobelins, he developed a considerable amount of work on color, including color classification, color applied to industry, as well as his most famous book on simultaneous contrast of colors, which had a great impact on several generations of artists as well as on color teaching. His exceptional longevity helped him to publish many books and hundreds of scientific papers, most of them on color topics. His 100th birthday was celebrated as a national event; he finally died at 103. His book *The Principles of Harmony and Contrast of Colors and Their Applications to the Arts* [1] was once considered one of the 12 most important books on color [2].

Chevreul's Life and Work

Born in Angers in 1786, Michel-Eugène Chevreul (Fig. 1) came to Paris when he was 17 years old and was appointed at the National Museum of Natural History as an assistant in charge of the chemical analysis of samples, thanks to a letter of recommendation from Vauquelin. Interestingly, his career as a chemist was determined by a sample of soap Vauquelin asked him to analyze. Indeed, the nature of animal fat was still unknown at the time. After several years of research, he published a book that gave him his fame as a chemist [3]. His discovery of the different acids contained in animal fat eventually led to important improvements in the field of industry, in particular in candles, as it made it possible to make candles shedding more light and less smoke. As he was also very interested in epistemological issues, he dedicated another book, 1 year later, to explain which method enabled him to make his discoveries [4].

In 1824, thanks to his fame as a chemist, Chevreul was appointed at the Gobelins Manufactures, as Director of the Dyeing Department. The Gobelins usually appointed chemists, as one of the main tasks of the Department of Dyes was

Georges Roque has retired.



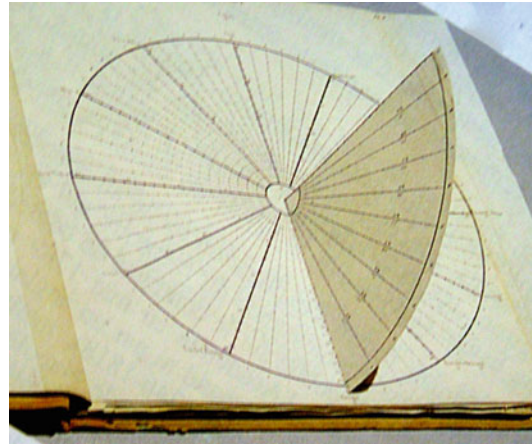
Chevreul, Michel-Eugène, Fig. 1 Michel-Eugène Chevreul (1786–1889) at the approximate age of 50. (Lithograph by Nicolas-Eustache Maurin, 1836, engraving by Conrad Cook)

to take care of the ► [dyes](#) of wools and silks used by the three manufactures, the most important being that of tapestries (Gobelins). The Director of the Department had indeed among other tasks that of looking after the quality of the wool (that had to be cleaned from grease and bleached) and the quality of the ► [dyes](#) according to their stability, their brilliance, and the kind of cloth to which they had to be applied (basically wool, silk, and cotton). Another important issue was that of color classification.

Before focusing this essay on the law of simultaneous contrast of color, it is worth noting the wide range of interests Chevreul had for colors. He himself suggested a classification of his work on color. The two main categories are (a) means of naming and defining colors and (b) ► [dyes](#) ([5], p. 121).

Means of Naming and Defining Colors

After being appointed at the Gobelins, Chevreul quickly realized that when the weavers needed a nuance of color, they used to show a sample of



Chevreul, Michel-Eugène, Fig. 2 Chevreul's chromatic hemispheric construction. (From Chevreul [1])

thread for matching, which was very empirical. For this reason, Chevreul felt it necessary to create a general classification of colors he first called “hemispheric construction” (Fig. 2), which, interestingly, is a black-and-white model (published in [1]). The circle is divided into 72 hues. Each of the 72 radii is divided into 20 segments, numbered from 1 to 20, corresponding to the scale of lightness, from the center (white) to the diameter (black). The third dimension is given by a quadrant perpendicular to the circle (unfolded in Fig. 2), corresponding to a saturation scale, and divided in 10 sections. This abstract system of color classification is the most complex realized at the time (1839) and permitted differentiating a great number of nuances: the 72 main hues of the circle with their 20 grades of lightness already give 1440 different nuances, to which must be added the 10 grades of saturation on the axe of the quadrant, that is, $1440 \times 9 = 12,960$. So the general amount of nuances is $14,400 + 20$ gray along the 10th radius, that is, 14,440. For a more precise account of Chevreul's color classification and a reply to the critiques made to it – in particular the fact that he would have confused lightness and saturation – see [6, 7], pp. 163–172.

Aware of the importance of color classification beyond the case of the Gobelins, Chevreul went on working on the topic and published several important books containing, unlike the 1839 black-and-white hemispheric construction,



Chevreul, Michel-Eugène, Fig. 3 First chromatic circle containing pure hues. (From M.-E. Chevreul, *Cercles chromatiques de M.-E. Chevreul*, 1861)

beautiful color plates [8–10]. At the 1851 *World Exposition* in London, Chevreul's chromatic circle was awarded a Great Medal. (Fig. 3) shows a further publication.

Dyes

In his classification of his own work on color, Chevreul divided the dye section into three parts: all that is relative to the simultaneous contrast of colors, all that concerns what he called the principle of color mixing (which corresponds to what is known today as chromatic assimilation), and finally chemical researches.

Indeed, long before being appointed by the Gobelins, Chevreul had worked on natural tints; on indigo, for instance, he devoted a dozen papers, the first being published in 1807, when he was 20 years old [11]. His interest for animal fat also helped him to work on the process of degreasing and of bleaching ► **dyes**, to which he devoted numerous papers.

Although Chevreul's work on color covers many aspects, his most important contribution to color is his law of simultaneous contrast of colors, as expounded in his book translated into English under the title *The Principles of Harmony and Contrast of Colours and Their Application to the*

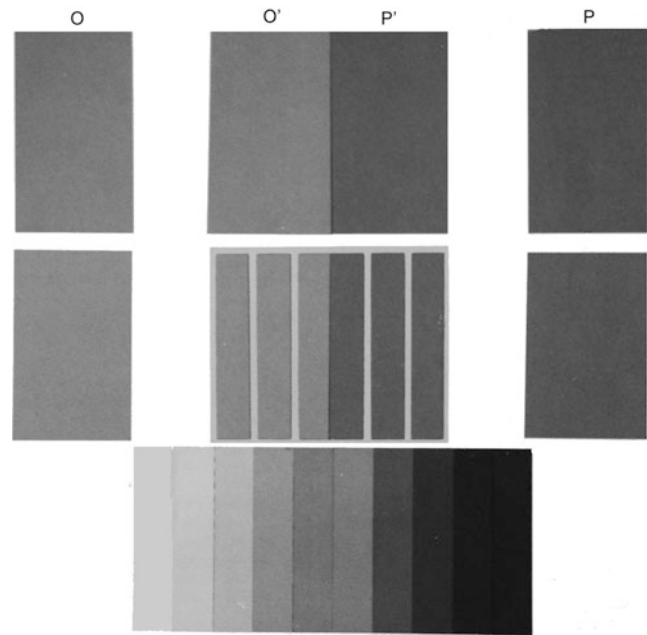
Arts (1st edition in French, 1839; 1st English translation, 1854). Its starting point was a complaint from the weavers of the Gobelins against the dyers of the Department of Dyeing that he directed. The complaint was in particular about the black samples of wool used for the shades of blue and violet draperies. As a chemist, Chevreul first tested the wools dyed in black in his workshop and compared them with those dyed in the best places from London and Vienna. After a careful comparison, he realized that the quality of the dyed material was not in question. This led him to raise a brilliant hypothesis: the lack of strength of the blacks was not due to the dyes or their uptake but was a visual phenomenon related to the colors juxtaposed to the blacks. This new discovery was all the more surprising as Chevreul, being a chemist, was not prepared to admit that the cause of the phenomenon he observed “is certainly at the same time physiological and psychological” ([12], p. 101).

Indeed, Chevreul realized that it is not the same to look at a sample of color when isolated and when juxtaposed to another contiguous one. In the latter case, the two samples appear different from when seen in isolation. This is the most general effect of the law of simultaneous contrast of colors that reads, “In the case where the eye sees at the same time two contiguous colours, they will appear as dissimilar as possible, both in their optical composition and in the strength of their colour” ([1], § 16). Chevreul made it very clear in one of the plates (Fig. 4): O and O', as well as P and P', have exactly the same degree of lightness; however, the perception of the samples differs when they are seen in isolation and juxtaposed to another sample of a different degree of lightness. The bottom of the same plate shows an effect known as “Chevreul's illusion”: each stripe (except the two extremes) being lighter than the following (when seen from left to right), a double effect is produced, because the left half of each stripe will appear darker and the right half lighter, due to the influence at the edges of the preceding and following stripes. (For Chevreul's illusion, see [13].)

Chevreul's demonstration is valid for hues as well as for lightness. For what concerns hues, the main definition of the law of contrast reads,

Chevreul, Michel-Eugène, Fig. 4

Illustration of the contrast of lightness. (Redrawn detail from original figure in Chevreul [1])



If we look simultaneously upon two stripes of different tones of the same colour, or upon two stripes of the same tone of different colours placed side by side, if the stripes are not too wide, the eye perceives certain modifications which in the first place influence the intensity of colour, and in second, the optical composition of the two juxtaposed colours respectively. Now as these modifications make the stripes appear different from what they really are, I give to them the name of *simultaneous contrast of colours*; and I call *contrast of tone* the modification in intensity of colour, and *contrast of colour* that which affects the optical composition of each juxtaposed colour. ([1], § 8)

The principle is exactly the same as for lightness: the modification consists in an exaggeration of difference. Yet, in the case of hues, what means exaggeration of difference? Chevreul's starting point is the concept of complementary colors, that is, colors considered as the most opposed. According to the knowledge of the time, Chevreul considered as complementary the following pairs of colors:

Red is complementary to Green, and vice versa;
Orange is complementary to Blue, and vice versa,

Greenish-Yellow is complementary to Violet, and vice versa

Indigo is complementary to Orange-Yellow, and vice versa. ([1], § 6)

So the modification perceived when seeing juxtaposed colors consists in perceiving each color as slightly tinted with the complementary color of the juxtaposed one. This is the clever way Chevreul had to understand and solve the problem raised by the weavers when complaining of the bad quality of the blacks dyed in the Dyeing Department of the Gobelins. When seen in isolation, the blacks are perfectly black, but when seen juxtaposed to violet, they are slightly tinted with the complementary color of violet, that is, yellow, and will look accordingly yellowish. In order to solve the problem, Chevreul suggested the weavers to mix a few threads of violet with the blacks, so that they neutralize the effect of yellow and make accordingly the blacks look blacker!

A particular case must be mentioned: what happens when the two juxtaposed hues are complementary, for example, red and green? According to the law of simultaneous contrast, the red will be slightly tinted by the complementary color of green, that is, red, and will be perceived as redder. Conversely, the green will be slightly tinted by the complementary color of red, that is, green, and will be perceived accordingly as greener. In this case, the two hues are not

modified in the sense of a transformation of the hue itself but enhanced.

Regarding the importance of Chevreul's law of ► **color contrast**, some authors hold that it was not original since other scientists before him, like Prieur, had already discovered the law of ► **color contrast** ([14], p. 306; [15], p. 140). It is true that Prieur and others had already discovered similar phenomena. However, it might be recalled that Chevreul fairly acknowledged what he borrowed from other authors (including Prieur) since he devoted a chapter to the issue of the relationship between his experiments and those made by others before him ([1], § 120–131). For him, indeed, his main contribution is *not* the “discovery” of the ► **color contrast** but the fact of classifying and structuring these phenomena well described by his predecessors but considered as belonging to one single class, when Chevreul proposed to carefully distinguish different kinds of contrast, so that the simultaneous contrast is just one of them. It is defined as follows:

In the *simultaneous contrast of colours* is included all the phenomena of modification which differently coloured objects appear to undergo in their physical composition and in the height of tone of their respective colours, when seen simultaneously. ([1], § 78)

Besides simultaneous contrast, Chevreul distinguishes *successive contrast*, which includes all the phenomena that are observed when the eyes, having looked at one or more colored objects for a certain length of time, perceive, upon turning them away, images of these objects offering the complementary color of that which is proper to each of them ([1], § 79). This distinction is very useful and helped to differentiate phenomena until then confused; it is still in use, even though simultaneous contrast is often related today to chromatic induction, while successive contrast is generally associated with chromatic adaptation; for this reason, the concept of afterimages is often used today instead of that of successive contrast.

Chevreul also distinguished a mixed contrast ([1], § 81), which combines simultaneous and successive contrast; it occurs, for instance, when, after having looked at one color for a certain length of time, another color is looked at. In this case, the resulting sensation is a combination of

the second color and of the complementary of the first one. Finally, Chevreul also added later a fourth contrast, the *rotary contrast* obtained with colored spinning disks [16].

It is out of the scope of this essay to discuss the main critiques addressed to Chevreul in particular the fact that he would have confused mixture of lights and mixture of pigments or simultaneous contrast and assimilation (for a full account of these issues, see ([7], pp. 93–102)).

Chevreul's Influence on Artists and Artisans

Another striking fact is the huge influence Chevreul had on generations of artists and artisans, even before the publication of his book on simultaneous contrast in 1839, thanks to the public lectures he gave and that were attended by painters, but also wallpaper fabricants and many other color practitioners. The range of his influence is indeed impressive, from tapestry to stained-glass restoration, shop signs to gardening. Many reasons explain the success of his book, soon printed out (the second French edition, published in 1889, as well as the third one, published in 1969, have also been quickly printed out). One is that by dedicating a copious volume to this matter, he gave wide public access to phenomena until then discussed only in specialized scientific journals. Another is that by meticulously studying the applications of his law to almost all the fields of art and crafts, he moved from pure science to applied science and addressed himself to almost all those who use color. One more reason is that by addressing the issue of how to match and harmonize juxtaposed colors, he provided artists and artisans with practical rules and harmony advices quite useful in the situation painters and tapestry-makers are constantly confronted with, that is, using juxtaposed colors. Finally, as he had a great prestige as a scientist, the color harmonies he proposed were avidly read and followed by artists anxious of matching their colors and enhancing them. Interestingly, unlike what is generally assumed ([17], p. 196), Chevreul was not a partisan of the harmony of color contrast and never recommended painters to juxtapose

complementary colors on their canvases. The reason is that for him the effect of simultaneous contrast always occurs naturally so that if a painter tries to imitate what he sees, he will exaggerate the effect instead of rendering it accurately.

Even though Chevreul's teachings gave rise to misunderstandings, he nevertheless had an enormous influence on painters, from the 1830s up to the beginning of abstract painting. If his influence on Delacroix remains controversial, it is important for ► [impressionism](#) and crucial for ► [neo-impressionism](#) and van Gogh. From the 1880s onward, his work was challenged by more up-to-date theories (Helmholtz, Rood); however, he still had an influence, in particular on Delaunay but also on color music, due to the usefulness of the rules of successive contrast he established. Even the most important books still used in color teaching (Itten and Albers) owe a lot to Chevreul. For a comprehensive account of Chevreul's influence on artists, see [7]. His huge importance on artists must not lead to overlook that his color theories also had a great influence on scientists, in a wide range of fields, including archeology, stained glass restoration [7, pp. 209–212], ophthalmology and anthropology [18].

Cross-References

- [Assimilation](#)
- [Afterimage](#)
- [Color Circle](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Complementary Colors](#)
- [Dye](#)
- [Impressionism](#)
- [Neo-Impressionism](#)
- [Simultaneous Color Contrast](#)

References

1. Chevreul, M.-E.: De la loi du contraste simultané des couleurs. ... Pitois-Levrault, Paris (1839). (The latest

- English translation is: The Principles of Harmony and Contrast of Colors and their Applications to the Arts (1855). Kessinger Publishing LLC, Whitefish, MT (2009))
2. Burchett, K.E.: Twelve books on color. *Color. Res. Appl.* **14**(2), 96–98 (1989)
3. Chevreul, M.-E.: Recherches chimiques sur les corps gras d'origine animale. F.-G. Levrault, Paris (1823)
4. Chevreul, M.-E.: Considérations générales sur l'analyse organique et sur ses applications. F.-G. Levrault, Paris (1824)
5. Chevreul, M.-E.: Recherches physico-chimiques sur la teinture. Comptes rendus des séances de l'Académie des Sciences. **X**, 121–124 (1840)
6. Heila, E.: The Chevreul color system. *Color. Res. Appl.* **16**(3), 198–201 (1991)
7. Roque, G.: Art et science de la couleur: Chevreul et les peintres, de Delacroix à l'abstraction, 2nd edn. Gallimard, Paris (2009)
8. Chevreul, M.-E.: Cercles chromatiques de M.-E. Chevreul, reproduits au moyen de la chromocalchographie. Digeon, Paris (1855)
9. Chevreul, M.-E.: Exposé d'un moyen de définir et de nommer les couleurs, d'après une méthode précise et expérimentale. ... Didot, Paris (1861)
10. Chevreul, M.-E.: Des couleurs et de leurs applications aux arts industriels. Baillière, Paris (1864)
11. Laissus, Y.: Un chimiste hors du commun: Michel-Eugène Chevreul. In: Sublime Indigo, exhib. cat, pp. 143–146. Musées de Marseille/Office du livre, Marseille (1987)
12. Chevreul, M.-E.: Sur la généralité de la loi du contraste simultané; Réponse de M. Chevreul aux observations de M. Plateau. ... Comptes rendus des séances de l'Académie des Sciences. **58**, 100–103 (1864)
13. Morrone, M.C., Burr, D.C., Ross, J.: Illusory brightness step in the Chevreul illusion. *Vis. Res.* **34**(12), 1567–1574 (1994)
14. Kemp, M.: The Science of Art. Optical Themes in Western Art from Brunelleschi to Seurat. Yale, New Haven (1992)
15. Mollon, J.D.: Chevreul et sa théorie de la vision dans le cadre du XIXe siècle. In: Roque, G., Bodo, B., Viénot, F. (eds.) Michel-Eugène Chevreul: un savant, des couleurs, pp. 137–146. Muséum National d'Histoire Naturelle/EREC, Paris (1997)
16. Chevreul, M.-E.: Compléments d'études sur la vision des couleurs. ... Firmin-Didot, Paris (1879)
17. Gage, J.: Chevreul between classicism and romanticism. In: Gage, J. Colour and Meaning: Art, Science and Symbolism, pp. 196–200. Thames and Hudson, London (1999)
18. Roque, G.: Chevreul's color-standard. In: Karliczek, A., Schwartz, A. (eds.) Farbe. Farbstandards in den frühen Wissenschaften, pp. 77–98. Friedrich-Schiller Universität, Institut für Geschichte und Medizin, Naturwissenschaft und Technik, Jena (2016)

Chiasma Opticum

► Optic Chiasm, Chiasmal Syndrome

Chinese Color Language

Victoria Bogushevskaya

Department of Humanities, University of Salento,
Lecce, Italy

Synonyms

Color categorization; Color idiom; Color lexicon;
Color naming; Color symbolism

Definition

Associative and symbolic meanings of culturally
the most significant color categories.

Notions on Color in Early China

In Western thought color has been associated with light (at least since the time of Aristotle), whereas in ancient China color was linked to the dichotomy colored/uncolored – either by nature, or artificially – and unrelated to the presence or absence of light. Colors were considered the property of the object itself and were therefore “perceived not as abstract concepts but as concrete substances, endowed with rich meanings” [1]. For instance, color correspondences applied to sacrificial animals during the Late Shāng 商 (ca. 1250–1046 BCE) period required that white, red, and multicolored animals were sacrificed in the ancestral cult; black sheep were used in the rain-making ritual; yellow animals were particularly addressed to the cosmic spirits of the earth and the cardinal directions [2]. Multicolored animals were inferior to those of pure color because they symbolized a ritual impurity. These first attempts of color categorization are registered in

Shāng oracle bone inscriptions (*jiǎgǔwén* 甲骨文), the earliest written records of Chinese civilization.

Color symbolism was directly related to the material from which the color was derived. Red was regarded as a “life-giving” color, a magical return of the soul: evidence from burial sites suggest that red ochre during Paleolithic and cinnabar from Neolithic times to the Hàn 漢 dynasty (206 BCE–220 CE) were interred along with the corpses [3]. Traces of cinnabar, applied with a writing brush, occur in the incised lines of writing on Shāng oracle bones.

Unknown in ancient Egyptian painting and the early Mesopotamian cultures, cinnabar, the naturally occurring mineral form of mercuric sulfide, was one of the most precious materials and significant colorants in ancient China. Finely ground and mixed with glue, it makes a brilliant and glowing red pigment. Another “magical” property of cinnabar, which is the color of blood and thus of life itself, is that it can be transformed into mercury, or quicksilver, a “living” metal. Mercury was also employed in early burial practices in China. The floor of the tomb of the First August Emperor of Qín 秦始皇帝 (r. 221–210 BCE) is said to contain a topographical model of the world with all the rivers running in mercury. Therefore, cinnabar had a profound religious significance in ancient China; it was believed to be a substance by which the forces of life and the cosmos could be controlled.

By the fourth century BCE, belief in the revitalizing function of cinnabar and mercury in burial practices had given rise to the idea of an elixir of immortality. Just as the living metal quicksilver symbolized the quickening of the soul, so the imperishable metal gold came to symbolize immortality. To attain immortality, the body must be transformed to a gold-like state. Early means of effecting this transformation involved consuming an herb of immortality or eating from vessels of gold. Perhaps because naval expeditions and golden dishes were out of the reach of most, there also arose the belief that gold and silver could be fabricated from other substances. The most important of these was cinnabar, solid and in its liquid state as mercury. This artificial gold could then be used to create the elixir of

immortality. Two fundamental beliefs of Daoist alchemy are, then, that the body can be transformed to a gold-like state, and that gold can be made from other substances, especially cinnabar. Cinnabar was so central to Daoist alchemy that the Chinese word for elixir, *dān* 丹, is the same word as cinnabar [4]. Present-day Chinese consider red auspicious.

One of the main characteristics of the psychology of an ordinary ancient individual is the concreteness of thinking associated with labor processes. In ancient China, these processes included pottery, bronzes, lacquerware, and textile manufacturing.

Implements of coloration were used already in the Neolithic period. Since the 1920s, Chinese and western archaeologists have discovered many sites of the Yǎngsháo 仰韶 “painted pottery” culture in central and northwest regions. The most distinctive wares with black-and-purple decoration were produced during the Bǎnshān 半山 phase of the Mǎjiāyáo 馬家窯 culture (Fig. 1).

The subterranean army of several thousand life-size terracotta soldiers, horses, and war chariots, under excavation since 1974 from the site near the supposed burial chamber of the First August Emperor of Qín, has become famous as one of the grandest archaeological discoveries of the twentieth century. Pictures of gray terracotta warriors, created to serve China’s first emperor in

the afterlife, have become known all over the world. Today it is the terracotta itself that determines the chromatic impression made by the excavated clay warriors, but originally the terracotta figures were colorfully painted (Fig. 2).

Red and green were the most frequent colors, followed by the numerous brownish and pinkish nuances ranging from reddish and violet-brown to pink tones and off-white. Inorganic materials of natural and synthetic origin were used as pigments. Remarkable is the generous use of pigments extracted from precious materials like cinnabar, malachite, and azurite. Before the pigment layers were applied, terracotta sculptures and their equipment (made of bamboo, wood or fabric) were coated with two very thin layers of East Asian *qī* 漆 lacquer, whereas bronze sculptures were painted directly without lacquer priming [5].

Lacquer is a natural product that comes from the sap of the lacquer tree. Out of 25 existing species of the lacquer tree, 15 grow in China. One of the earliest excavated lacquered samples is a wooden bowl with red cinnabar lacquer of Hémùdù 河姆渡 culture was found in Yúyáo 余姚, Zhèjiāng 浙江 (ca. 6000–ca. 5000 BCE) [1]. Lacquer was often applied on objects made of wood, leather, bone, and textiles.

As is commonly known, China is the birthplace of sericulture; silk was used there as early as six thousand years ago. Other textiles existed, such as ramie produced from nettles, hemp from mulberry, abutilon from mallow, and some may even predate silk [6]. The art of dyeing naturally developed along with the creation of textiles. The surviving examples of the Warring States 戰國 period (475–221 BCE) textiles reveal a high-level dyeing technique of vegetable and mineral dyes, decorated with painting and embroidery (Fig. 3). They are dyed in such a variety of red, yellow, green, blue, and purple shades that the color lexicon of that time was unable to convey all the nuances of the obtained hues.

Unsurprisingly, the modern Chinese word for “coloration” includes the morpheme (*cǎi* 彩 < 採) that was originally applied to embroidery (when the yarns were selected according to their color) and implied the meaning “combined by interwoven multicolored threads.” As regards the morpheme “color” (*sè* 色), it originally referred to



Chinese Color Language, Fig. 1 Funerary urns. Bǎnshān type red earthenware with black and purple painted decoration. The bold linear decoration is executed with a soft brush. Production date ca. 2650–2350 BCE. Ht.: 34–36 cm. (© The Trustees of the British Museum)



Chinese Color Language, Fig. 2 Left: General T20G10:97, pit no. 2. Right: its plaster cast painted reconstruction [4]

“facial expression,” “complexion,” “outward appearance,” and is used to describe naturally or artificially colored objects in China’s first etymological dictionary, the *Shuōwén jiězì* 說文解字 of Xǔ Shèn 許慎 (ca. 100 CE).

“Primary” and “Secondary” Colors

NB: It should be immediately warned that:

- (a) Colors described hereinafter do not always coincide with the coloristic meanings accepted in European art.
- (b) The meanings of many color names have shifted. “Each individual shift, when it occurs, has a clear connection with the older meaning, if further shifts occur, they are increasingly likely to obscure or supersede the earlier meanings” [7]. The definitions of color names used here are those of Modern



Chinese Color Language, Fig. 3 Silk textiles from a tomb at Mǎshān 馬山, sixth to third century BCE. Left: detail of the embroidered sleeve of a garment in the sixth of

thirteen layers of wrappings and clothing. Right: detail of embroidered phoenix decoration on the quilt or cover forming the fourth layer [5]

Standard Mandarin (henceforth MSM), that is, contemporary Chinese from the twentieth century onwards.

The first proper color categorization one can see in the literature of the Eastern Zhōu – Hàn period, which is full of references to the “five colors” (*wǔ sè* 五色): “white” *bái* 白, “red” *chì* 赤, “yellow” *huáng* 黃, “grue” (green-and/or-blue) *qīng* 青, and “black” *hēi* 黑. The denotations of these terms are difficult to trace [8].

The texts referring to the five colors were probably influenced by the *wǔ xíng* 五行 theory, conventionally translated as the theory of five elements, or the theory of five agents, or which assumed complex interrelationships among the five elements, wood (Wd), fire (F), soil (S), metal (M), and water (W), the five directions, the five colors, the five tastes, the five smells, the five internal organs of the body, etc. (see Table 1). The pentamorous structure is often seen as the 2 + 3

formula that presumably represents the *yīn-yáng* 陰陽 duality and the triad of Chinese thought, Heaven-Earth-Humanity (天、地、人). Thus, the Chinese of that time “had theoretical reasons for believing that there *had* to be five colors, and their statements to this effect cannot be taken as direct evidence about their language (they were, after all, capable of tacking on a fifth direction, ‘middle’, to fit to the theory)” [9].

The correspondences shown in Table 1 are only a few out of many more; they were the commonplaces of thought in growing measure from the mid-third century BCE onwards and may be found in varying degrees of completeness in most of the ancient texts. They are full of discrepancies and may be stated in different ways in the same text.

Remarkably, the existence of the five-color gamut in ancient China is also confirmed by the archaeological discoveries. In the 1960s there were found 15 vessels *túlú* 圖盧, color-containers

Chinese Color Language, Table 1 Pentamorous symbolic correlations

Element	Wood (Wd)	Fire (F)	Soil (S)	Metal (M)	Water (W)
Season	Spring	Summer	Late summer	Autumn	Winter
Direction	East	South	Middle	West	North
Color	Green/blue	Red	Yellow	White	Black
Directional animal	Grue Dragon	Vermilion Bird	Yellow Dragon	White Tiger	Dark Warrior
Human psycho-physical function	Vision	Thought	Speech	Hearing	Demeanor
Source of happiness	Physical and mental health	Fulfilling to the end the will of heaven	Richness	Love of virtue	Longevity
Taste	Sour	Bitter	Sweet	Acrid	Salt
Smell	Goatish	Burning	Fragrant	Rank	Rotten
Planet	Jupiter	Mars	Saturn	Venus	Mercury
Weather	Wind	Heat	Thunder	Cold	Rain
Animal category	Scaly	Feathered	Naked	Hairy	Shell-covered
Viscera	Spleen	Lungs	Heart	Kidneys	Liver
Affective state	Anger	Joy	Desire	Sorrow	Fear

assigned to the Late Shāng – Western Zhōu 西周 period. Most of these containers are made in bronze and vary from cuboid to round and triangular in shape; each of them has three to five tubular receptacles for the pigments. In five of these containers, residues of pigments were found in the bottom of the tubes, and these have been identified as white, black, red, green, and yellow powders. A small scraping from the bottom of the four receptacles of one of the containers underwent spectroscopic analysis, and the results showed that the following substances were present in a large amount: white powder, calcium oxide or calcium carbonate; black powder, carbon black or graphite; red powder, iron oxide; green powder of compound of copper [10].

Five elements constantly interact in some 20 combinations and permutations affect the cosmos, the natural world, and the world of human affairs. The most culturally significant arrangements are the “mutual production” (*xiāng shēng* 相生) and the “mutual conquest” (*xiāng kè* 相剋): the former arrangement (WdFSMW) was applied to illustrate life cycle, while the latter arrangement (WdSWFM) was extended to dynastic transitions. If a ruling dynasty’s emblem was fire, the honored direction was the south and the worshipped color

was red, the overcoming dynasty’s emblem would become water, therefore the honored direction would be the north and the worshipped color would be black. As a whole, the system is bound up with Chinese astronomy, astrology, geomancy, and also with a branch of ancient Chinese medicine.

Today in Beijing, in the entrance part of the Forbidden City (*Zijincheng* 紫禁城), within Zhōngshān park 中山公園, one can see the Altar to the Gods of soil and grain (*Shèjītán* 社稷壇). Built in 1421, it was used for performing sacrifice ceremonies by the emperors of the last two dynasties that ruled China. The altar is a small, three-tiered square platform: according to the ancient Chinese mytho-geographical and cosmographic principle, “heaven is round and the earth is square” (*tiān yuán dì fāng* 天圓地方). The surface of the Altar is paved with five types of soils, distributed into sections according to the appropriate cardinal directions (Fig. 4). The soil in five colors symbolizes the traditional idea that “all under heaven belongs to the emperor” (*pǔ tiān zhī xià, mò fēi wáng tǔ* 溥天之下, 莫非王土) and the everlasting endurance of the dynastic world.

In Confucian writings, the colors of the five elements are orthodox or conventional and

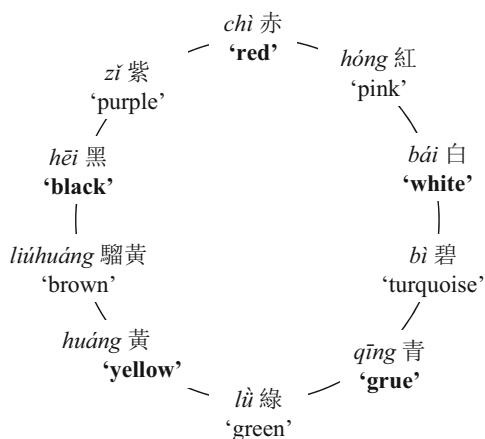


Chinese Color Language, Fig. 4 The altar of five-colored soil (五色土) in Beijing. (Photo by Victoria Bogushevskaya)

referred to as “primary,” “correct” (*zhèng* 正), whereas the colors generated in the process of the mutual conquest of the respective elements are called “intermediate” (*jiàn* 間) and are considered subordinate. The original meanings of the “secondary” colors are “reddish-pink” *hóng* 紅, “yellowish-green” *lǜ* 綠, “purple-red” *zǐ* 紫, “turquoise” *bì* 碧, and “brown” *liúhuáng* 騮黃. The first three color names are textile terms, the penultimate one is semantically related to jade and the latter one to horse [8] (Fig. 5).

Technically, there was such a great variety of the intermediate tints at that time, that any shade lighter or darker than the “standard pure” color could be called “intermediate,” and therefore automatically considered secondary and of a low rank. The five “primary” colors were advocated by the *Yīn-Yáng* 陰陽 scholars during the Warring States period and later by Confucians, as they simply had to adapt contemporary advanced dyeing knowledge to their theory, and the list of the so-called secondary colors inevitably appeared.

It is important to note that the word for “red” (*chì* 赤) in the five-agent theory is different from the word for “cinnabar” (*dān* 丹). Alchemy was condemned by the Confucian scholars at court as early as the first century BCE. From a Confucian point of view, alchemy was at best mere fakery and illusion, and at worst, a waste of money, talent, and lives. The Confucianists had no interest in the physical manipulation of pigments; their



Chinese Color Language, Fig. 5 Canonical colors' mixing order [8]

approach to color instead made symbolic and moralizing use of hue, that is, “the attribute of colors that permits them to be classed.” Color as hue is charged with social and political significance. In contrast with Daoist notions of color as a controlling substance, the Confucian belief is that color requires control [4]. The distinguishing between “primary” and “mixed” colors inspired “noble versus humble,” “good versus evil,” and other symbolic meanings.

It is interesting to note that the names of the dress decorations used by ancient painters and embroiders depend on the combinations of five colors: *wén* 文 is a decoration which is composed of blue-green and red, *zhāng* 章 – of red and white, *fū* 黼 is an axe-pattern decoration with interlacing black and white colors, *fú* 黻 is a 亞-pattern decoration with interlaced black and blue-green colors, *xiù* 繡 is an embroidery decoration mixed with five colors (blue-green, red, white, black, and yellow) [11].

As regards the descriptions of obtained shades, ancient sources often provide very limited information. We can read descriptions like:

Three dips produce [what is called] *xūn* 纁; five dips make [what is called] *zōu* 緅; and seven dips result in [what is called] *zī* 緇 [11].

One, therefore, needs to investigate what pigment or dye was used, with or without mordants.

Color became an indispensable element of official decrees on clothing in imperial China (from the Qín 秦 dynasty, 221–207 BCE, to the Qīng 清 dynasty, 1644–1911) to distinguish ranks and contribute to the stability of social hierarchy and bureaucratic systems. Each dynasty regulated pattern, color, material, and the number of patterns and accessories used for coronets, costumes for various occasions, and accessories that people of different ranks could wear [12]. The “primaries” were limited to higher-ranking officials, for upper clothing, and headgear, while the “secondaries” must be used by those of lower rank and for clothing the lower body [6]. There were also rules dictating the correct clothing for each season and when these seasonal wardrobe changes should be made.

As regards the concrete dyestuffs, a Táng-dynasty (618–907) text is more specific. The *Táng liùdiǎn* 唐六典 (Compendium of administrative law of the six divisions of the Táng bureaucracy) states that court dyers recognized five official colors: red from madder, blue/green from indigo, yellow from gardenia, “black” from acorns, and purple from gromwell. Except for acorn’s gray-brown-black range, these were the colors used to dye official robes the appropriate rank colors and, with the exception of purple, they were the five colors associated with *wǔ xíng* 五行 [13].

During the golden age of alchemy, from the third century BCE to the ninth century CE, painting was dominated by the mineral pigments. Several connections existed between alchemy and painting in mineral colors. Many of the same mineral substances served as ingredients in the elixir of immortality and as painting pigments. These include cinnabar, vermilion, malachite, azurite, realgar, minium, orpiment, ochre, gold, silver, lead, graphite, clamshell, and alum. Many of the scientific techniques used to process these materials for use in alchemy and painting were the same, such as grinding, heating, washing, and suspending. Moreover, alchemy and painting were both referred to by synecdochical names. While alchemy was called “the art of yellow and white” (*huáng bái zhī shù* 黃白之術), in reference to the desire to produce gold and silver, painting

was called the art of “red and blue,” or literally, “cinnabar and azurite” (*dānqīng* 丹青). From the Daoist point of view, both painting and alchemy were transformative and magical arts that could be used to control the forces of life [4].

The color grading system of the last two dynasties to rule China, the Míng 明 (1368–1644) and Qīng 清 (1644–1911), involved more grades and colors than those of previous dynasties. Color restrictions were applied to the emperor’s attire for particular occasions and also to the use of colors for clothing and utensils appropriate to various degrees of status within the imperial household. For example, in the formal court attire used only on ceremonial occasions, the emperor would wear blue for sacrifices at the Altar of Heaven, red at the Altar of the Sun, light blue for sacrifices to the moon, and yellow for those to the earth. Yellow was also the default color for ceremonies where none was stipulated and was furthermore the color associated with the Qīng dynasty. Yellow and green porcelain glazes were used in different combinations on bowls and dishes according to the imperial rank of the user [6]. Color symbolism has been omnipresent at the palaces and temples.

The last emperor Pǔ-yí 溥儀 (1905–1967) of the Qīng dynasty describes the dominance of the imperial yellow in the following way: “The glazed tiles were yellow, my sedan-chair was yellow, my chair cushions were yellow, the linings of my hats and clothes were yellow, the girdle round my waist was yellow, the dishes and bowls from which I ate and drank, the padded cover of the rice-gruel saucepan, the material in which my books were wrapped, the window curtains, the bridle of my horse... everything was yellow. This color, the so-called ‘brilliant yellow’, was used exclusively by the imperial household...” [14].

Purple, originally considered to be a “secondary” color, was nevertheless added to the palette of colors to indicate very high rank presumably in the second century BCE, although not without protest from conservative Confucian scholars. Extracted from the roots of a gromwell plant, *Lithospermum erythrorhizon* (*zǐcǎo* 苺草/紫草), reddish-purple had been one of the most popular colors until the Former Hàn period

(206 BCE–9 CE) [13]. The ancient states of Lái 萊 and Qí 齊 were famous for the production of this plant [8]. The gromwell-dyed silk was costly, and the dye process required considerable skills. The color was derided by Confucian scholars as example of the luxury and frivolity that they deplored [13]. Furthermore, Confucius was original of the state of Lǔ 魯, a long-time enemy of the states of Lái and Qí. Another possible reason is that purple was the color associated with Daoism (see section “Purple”).

A pigment only known in China is the bright bluish-purple, a man-made barium copper silicate, which was widely used until the end of the Hàn dynasty (220 CE). Rediscovered in 1983, it is known as Chinese Purple or Hàn Purple (*Hàn zǐ* 漢紫) today and is one of the most frequent colors of the Qín Shǐ Huángdì’s terracotta army. It is not known if artificial barium copper pigments were produced before the Qín dynasty [15].

Daoism was a strictly Chinese religion which had originated and flourished entirely on Chinese soil. The robes of the Daoist priests were rather drab and plain; but for grand occasions, such as special funeral rites, the high priests wore richly decorated copes. By contrast, Buddhism was a foreign religion in China, introduced from India by way of Central Asia; as such, during the Middle Ages it had even been proscribed for a time, along with Manichaeism, Nestorian Christianity, and other foreign faiths [16]. With time, Buddhism absorbed many Chinese cultural elements and became Chinese Buddhism.

Just as a monk’s robe *jiāshā* 袈裟 (from the Sanskrit *kāṣāya*) distinguished him from an official, variations in the robe also distinguished different types of Buddhist monks. A common term for monks in medieval texts is the “black-robed ones” (*zīyī* 緇衣). And when referring to monks and laymen, Buddhist texts commonly use the expression “the black and the white” (*zībái* 緇白) [17]. The historical Buddha had cautioned his followers against luxurious living. Accordingly, when his monks were presented with valuable silks or discarded robes, they would simply cut these up into small squares and reassemble them without regard to pattern, to fulfill the spirit if not the strict letter of the law [16]. In medieval China, monks

from different regions were recognized by the color of their robes – pitch-black (*hēi* 黑) in the Jiāngnán 江南 region, brown (*hè* 褐) in the area around the capital at Kāifēng 開封, and so on. After the tenth century CE, the color of robes was gradually standardized. Now Chinese Buddhist monks usually wear black or gray robes [17].

Chinese Mnemonic Phrases for the Colors of a Rainbow

Many languages have a mnemonic phrase to help remember the order of the seven colors of the rainbow. We know the English “Richard-of-York-gave-battle-in-vain” or the Russian “Каждый (красный ‘red’) охотник (оранжевый ‘orange’) желает (жёлтый ‘yellow’) знать (зелёный ‘green’), где (голубой ‘light blue’) сидит (синий ‘blue’) фазан (фиолетовый ‘purple’)” [“Every hunter wants to know where the pheasant’s hiding”], where the initial letters of each word correspond to the rainbow color sequence.

In Chinese, there are vague analogs. The first one is related to the so-called “folk” version of the five elements’ sequence, “*jīn-mù-shuǐ-huǒ-tǔ*” 金木水火土 (MWdWFS). The phrase has the 2 + 3 prosodic structure and is therefore easy to memorize. As a result, the correlative colors automatically emerge in a speaker’s mind.

In 1853, Zhāng Fúxī 張福僊 translated *Guāng lùn* 光論 (*On light*), which was the earliest systematic study on chromatic dispersion in China [18]. However, the color terms he used never crystallized into a mnemonic phrase. Instead, the phrase that is taught at school today derives from the line of a poem written by Máo Zédōng 毛澤東 in the 1930s: “*chì-chéng-huáng-lǜ-qīng-lán-zǐ*” 赤橙黃綠青藍紫 “red, orange, yellow, green, blue, blue, purple” [19].

Symbolism of the Main Color Categories

Studies on Chinese color terminology are carried out from different perspectives (historical-cultural, participation of a term in

word-formation, psycholinguistic analysis, usage frequency, etc.) and often aim to verify the evolutionary stage encountered in contemporary color vocabulary along with the list of the so-called basic color terms. These issues need not be addressed here. For the convenience of comparison with other languages, this section aims to describe symbolic and figurative meanings of the most commonly used color categories in MSM. Some of these meanings can be cross-culturally similar, intuitively graspable or remote.

Each color category includes a wide range of shades. In addition to its cosmological symbolism (see Table 1), each category as well as the shades within the same category may have several different and sometimes mutually exclusive meanings. These meanings often originate from popular beliefs and philosophic systems.

Black

As a symbol, black stands for darkness, death, honor. “Dark-cool” *xuán* 玄 appears in old mythological and astronomical traditions. The mythical ancestor of the Shāng was a *xuánniǎo* 玄鳥 “black (dark-cool) bird.” “Although the north was associated with *hēi* 黑, the association with *xuán* 玄 may be older: the northern constellations (also regarded as a god) are called *xuánwǔ* 玄武 the ‘dark warrior’” [9].

In Buddhism, black symbolized “bad” in a sense of violating the principles of Buddhism. In the old Chinese calendar, unlucky days are black (*hēidàorì* 黑道日) [20].

In traditional theatre plays, the choice of color in face painting often derives from the written literature about the historical characters who appear on the stage. Bāo Zhēng 包拯, the famous judge from the Northern Sòng 北宋 dynasty (960–1127), was called “iron face” because he was impervious to bribery and influence. His stage counterpart, Judge Bāo, is now depicted with black and brown face paint [21].

If a woman’s hair is described as *wūyún* 烏雲 “raven-black clouds” this means that it is black and beautifully lustrous.

One of the five forms of ancient corporal punishment (the other four include cutting off the nose, cutting off the feet, castration, and

decapitation) is called *qíng* 黥 (the character contains the semantic determiner “black” 黑), “ink punishment,” in which the forehead was cut and marked with a black dye as a permanent mark [18].

In MSM, black has negative connotations. Examples include *hēixīn* 黑心 (“black heart”) “evil mind,” *hēishèhuì* 黑社會 (“black society”) “criminal underworld,” *hēihuò* 黑貨 (“black goods”) “contraband,” *hēikè* 黑客 (“black guest”) “hacker,” “cyber-terrorist,” *hēichē* 黑車 (“black vehicle”) “illegal taxi,” *hēiháizi* 黑孩子 (“black child”) “unregistered child” (a child whose birth means the birth quota has been exceeded), etc.

White

“White tiger” *báihǔ* 白虎 is the animal of the West. White is said to have been the symbolic color of the Shāng. In the seventh century, scholars who had not yet been appointed to posts wore white clothes. The God of Earth *Tùdishén* 土地神, worshipped in folk religion, is represented as an old man with a white face [3]. White was adopted by Buddhism to mean “good” [14].

On the Chinese stage, actors with white faces represent men who are cunning and treacherous. Cáo Cāo 曹操, the ruthless warlord in the Three Kingdoms stories, has the white face painted with cake makeup to conceal his deceitful nature. The purity of the white symbolically hides his evil intentions (Fig. 6) [21]. The idiom *chàng báiliǎn* 唱白臉 (“play the white-face”) means “to play the role of the villain.”

One should never wear anything white in one’s hair, as this is very unlucky. The west is not only the direction where the sun dies, but also where dead people are headed for. Funeral is *báishì* 白事 “white affair.” However, it is not quite correct to say that mourning is white *bái* 白. Mourning apparel was made of *sù* 素, a kind of unbleached sackcloth, which is a brownish-yellow rather than white [3].

“Pure white” *qīngbái* 清白 is an expression denoting a morally pure, stainless person. Hair color of an elderly person is *bái* 白 white, but a blonde person’s hair is *jīn* 金 golden or *huáng* 黃 yellow. “To give someone a white eye” (*gěi báiyǎn*



Chinese Color Language, Fig. 6 *Cáo Cǎo*, the ruthless warlord in the Three Kingdoms story, has the white face to conceal his deceitful nature [15]

給白眼), that is, to look askance and show the whites of one's eyes, is a gesture of disdain, disapproval or anger.

In some cases, *bái* 白 has nothing to do with color at all; the meaning may imply “in vain,” as in *báizuò* 白做 “waste one's energy,” “plain,” as in *báikāishuǐ* 白開水 “[plain] boiled water”, or “empty,” as in *báichī* 白痴 “idiot,” “imbecile.”

Red

This category used to be represented by the term *chì* 赤 that denoted a wide range of red hues, from dark red to dark orange and reddish-brown. Among other terms for red, *zhū* 朱 was also common enough to be a possible rival of *chì* 赤. From the *Táng* 唐 dynasty (618–907) onwards, *chì* 赤 was gradually ousted by *hóng* 紅 (originally one of the “secondary” colors with the meaning “pink”) and is a basic color term for RED in MSM

[22]. Nevertheless, *chì* 赤 continues to be used as an epithet of things in which the red color forms a natural or obvious mark of kind or class. The choice of *chì* 赤 or *hóng* 紅 in various expressions is of totally etymology-related lexical nature [14].

Related to the element “fire” and the color red *chì* 赤, South is the quintessence of the *yáng* 陽 male natural principle. The throne where the Emperor sat was always facing the South.

In Buddhism, “red dust” *hóngchén* 紅塵 means “the world of mortals” [20].

Hóng 紅 originally was a “secondary” non-*yáng* and therefore “womanly” “reddish-pink” color: *hónglèi* 紅淚 (“rosy teardrop”) “women's tear,” *hóngyán* 紅顏 (“rosy cheeks”) “a beauty.”

NB: “Red chamber” *hónglóu* 紅樓 means “women's living quarters” (the sheltered chambers where the daughters of prominent families resided and where a stranger was not allowed to enter), whereas “cinnabar-red gates” *zhūmén* 朱門 refers to wealthy homes.

Red is a life-giving and an auspicious color (see section “Notions on Color in Early China”). “Vermilion bird” *zhūquè* 朱雀 is the directional animal of the South. A red raven *chìwū* 赤烏 was the symbolic creature of the *Zhōu* dynasty, because the *Zhōu* identified themselves with the sun [3].

On the Chinese stage, a red-faced man is a holy person or a loyal general. *Guān Yǔ* 關羽, from the Three Kingdoms story, was referred to as having a “red ochre complexion,” and his stage makeup is a plain red *hóng* 紅 face [21]. The idiom *chàng hóngliǎn* 唱紅臉 (“play the red-face role”) means “to play the good cop.”

Red *hóng* 紅 means “successful,” as in *hóngyùn* 紅運 (“red fate”) “good luck” or “happiness,” as in *hóngshì* 紅事 (“red affair”) “wedding.” Place for a wedding ceremony is to be decorated with red. However, the bridal clothes, coverings for the marriage altar, and ornamental curtains, in use until about 1900, often presented a relatively softer red, which often faded out into a light golden tone, having been made of local vegetable dyes that were difficult to fix firmly [16].

Red is also held to be the color of wealth. In MSM, it has the meaning “profit,” as in *fēnhóng*

分紅 (“divide red”) “share out bonus,” usually distributed in *hóngbāo* 紅包 red envelopes.

Of the three gods of good fortune, highly regarded in Chinese folk religion, collectively known as “*Fú-Lù-Shòu sān xīng*” 福祿壽三星 (three stellar deities of happiness, wealth, and longevity), the one who confers high office and happiness wears a red robe; the one dressed in green blesses a family with children and wealth, while the third, who gives long life, is dressed in yellow or in white (Fig. 7).

Since red and green are the colors of life, their combination is especially significant. In one novel we read that a girl is as red as peach blossom and ripe for love (*táo hóng* 桃紅), while others are as green as willows (*liǔ lǜ* 柳綠) and can only be pitied (“pity” is a word that men often use when they are describing their feelings for young women). The idiom “lanterns red, wine green” (*dēng hóng jiǔ lǜ* 燈紅酒綠) describes the way of life in low pubs [3].

In MSM, red *chì* 赤 can also mean “bared,” as in *chìzú* 赤足 (“*chì* + foot”) “barefooted,” or “sincere,” as in *chìxīn* 赤心 (“*chì* + heart”) “sincerity.”

Red is associated with revolution and communism. With certain Chinese expressions, the word *red* does not help to convey the meaning in English. For example, *yòu hóng yòu zhuān* 又紅又專 “both red and expert” is not as helpful as

“both socialist-minded and professionally qualified.”

A common term for convicts in imperial China is the “red-ochre-robed ones” (*zhěyī* 赭衣). Convict’s garb was of reddish-brown color, collarless and with raw edges [18].

NB: Black tea is “red tea” *hóngchá* 紅茶 for Chinese, as the reference is the reddish-brown color the brewed tea liquid obtains, whereas Europeans refer to the color of the dry tea leaves. Brown sugar is *hóng* 紅 red or *chì* 赤 red [22]. If envious, one becomes *yǎnhóng* 眼紅 “red-eyed.”

Yellow

The coupling of “yellow” *huáng* 黃 and “middle” *zhōng* 中 (China is the “Middle Kingdom,” *Zhōngguó* 中國) is no doubt connected with the annual deposit of loess from the Gobi on the North Chinese plains, which then turn yellow; and this too was the color of the Emperor who ruled the Middle Kingdom. On the one hand, it was thought that the emperor was located in the center of the five directions. On the other hand, yellow was considered as the color of the sun: just as there cannot be two suns in the sky, there cannot be two emperors in a nation [12].

“Yellow Emperor” *Huáng-dì* 黃帝 was the mythical Emperor, but it was not until 618 CE that reddish-yellow *chihuáng* 赤黃 was accepted



Chinese Color Language, Fig. 7 From left to right: God of Wealth (nineteenth century Ht.: 24.40 cm), God of Happiness (eighteenth century Ht.: 25.40 cm), God of

Longevity (eighteenth century Ht.: 10.5 inches). (© The Trustees of the British Museum)

as the fitting color for the Imperial Majesty; hitherto red had been the preferred color [3].

The severe regulations for the costume of the Qīng dynasty that were promulgated in 1759 and enforced in 1766 – described, with pictures, in the *Huángcháo lǐqì túshì* 皇朝禮器圖式 (illustrated precedents for the ritual paraphernalia of the Qīng court) – required that the emperor and empress should wear robes of pure clear yellow *míng huáng* 明黃 “bright yellow” for any special occasions (Fig. 8), except for certain great annual sacrifices when the emperor’s dragon robe had to be of a special color appropriate to the particular occasion (see section ““Primary” and “Secondary” Colors”). Yellow robe with lack lower sleeves and outer jacket could be black for more formal occasions [16]. The court robe and *mǎng* robes (*mǎng páo* 蟒袍) for high princes should be *xīng huáng* 杏黃 “apricot yellow” (Fig. 9), those of lesser ones be *jīn huáng* 金黃 “golden yellow” (Fig. 10), while imperial family members of lower ranks were permitted to be *lán* 藍 blue and *shí qīng* 石青 “azurite blue” [12]. No one outside the imperial family could use any kind of yellow unless the privilege was expressly conferred.

Yellow is taken to symbolize fame, progress, and advancement. In the old Chinese calendar, lucky days are yellow (*huáng dào* (jì)rì 黃道(吉)日), which possibly stems from the Eternal Wheel of life in Buddhism [20].

In popular beliefs, yellow is associated with death. This probably stems from the correlation

of the color yellow with the element “soil.” The “Yellow Springs” (*huáng quán* 黃泉) is a term denoting the underworld. The expression “to return to the Yellow Springs” (*mìng guī huáng quán* 命歸黃泉) means “to meet one’s end.” In Daoism, yellow is the symbol of the deified Lǎo zǐ 老子 (the ancient philosopher of the Spring and Autumn 春秋 period (770–476 BCE) who was regarded by the Daoists as the founder of their school). In landscape painting and poetry, yellow is related to autumn and faded leaves and, by extension, to the end of the year and humans’ advanced age. It seems plausible that the association between dropping leaves with failure of a plan and disillusionment with life inspired the negative meaning of yellow in MSM, which is “failure to achieve something” [14].

Yellow also means “infant” and derives from the compound *huáng kǒu* 黃口 “yellow mouth,” which originally denoted “nestling” as the mouths of baby birds usually appear yellow [14]. On the Chinese stage, a yellow face is designated for devious and shrewd characters [21].

In imperial China, the term *huáng mén* 黃門 (lit. “yellow gate”) referred to eunuchs. In MSM, any form of the arts or literature which implicitly or explicitly touches upon the topic of sex or



Chinese Color Language, Fig. 8 Embroidered yellow silk *lóng páo* 龍袍 dragon robe (made ca. 1800). Ht.: 144 cm. (© Chester Beatty Digital Collections)



Chinese Color Language, Fig. 9 Women’s apricot yellow silk court robe (made ca. 1900). Ht.: 146 cm. (© Chester Beatty Digital Collections)



Chinese Color Language, Fig. 10 Man's golden yellow embroidered silk court robe (made ca. 1800). Ht.: 135 cm. (@ The Palace Museum, photographed by Hui Feng [18])

nudity can be labeled as “yellow.” The compounds *dǎhuáng* 打黃 “to attack the yellow,” *sǎohuáng* 掃黃 “to wipe out the yellow” refer to the campaigns against pornography. On the other hand, a “yellow-blossom woman” (*huánghuānǚ* 黃花女) is one who is still a virgin.

NB: “yellow movie” means “pornographic movie,” whereas the famous telephone directory, the *Yellow Pages*, is not a “yellow book” in the Chinese sense.

Green and Blue

Qīng 青, one of the canonical colors, covered all shades from green and blue through dark-gray. The standard combinations include *qīngtiān* 青天 “[blue] sky” and *qīngcǎo* 青草 “[green] grass.” In MSM, there are separate, psychologically salient terms for green (*lǜ* 綠), blue (*lán* 藍), and gray (*huī* 灰). *Qīng* 青, however, continues to function as a color term with a context-dependent meaning [24].

“Rising straight up in a blue sky” (*qīng yún zhí shàng* 青雲直上) is a metaphor for a meteoric rise in one’s career. The “blue bird” *qīngniǎo* 青鳥 is a messenger from the mythical Queen Mother of the West; hence, any messenger. However, one should never wear light or dark blue flowers or ribbons in the hair: it is unlucky. Traditionally, blue eyes are regarded as ugly [3].

The word *lán* 藍, the basic color term for BLUE in MSM, derives from the name of the indigo plant, the phrase “Green-blue is obtained from the indigo plant, but is more green-blue than the indigo plant” *Qīng qǔ zhī yú lán ér qīng yú lán* 青取之於藍而青於藍 figuratively means “The student surpasses the master.”

Kuí-xīng 奎星, the god of literature, was originally a scholar, who was frustrated in his ambitions and committed suicide, is often represented with a blue face and wears green robes [3]. On the Chinese stage, blue face indicates an arrogant rascal [21].

The figurative meaning of *qīng* 青 is “not too far advanced in growth or life” or even “careless.” Thus “green years” *qīngnián* 青年 (“green” + “years”) means “youth,” *lèngtóuérqīng* 愣頭兒青 (“reckless” + “head” + “green”) refers to a quick-tempered person and *qīngpí* 青皮 (“green” + “peel”) stands for “an idle mischievous person.” Internet slang includes the terms like *qīngwā* 青蛙 “[green] frog” that refers to an ugly young person and 小青馬 *xiǎoqīngmǎ* “small green horse” that means “a young Marxist.”

Green dyes had always been most difficult for the Chinese artisans. Indeed, in former times green was a problem for dyers everywhere. In spite of the prevalence of this color in nature, it proved almost impossible to obtain a successful green dye. Only the berries from buckthorn, or certain other shrubs from the *Rhamnus* family, could produce one, but they could not be collected in any practical quantity. Therefore, the common practice was to dye threads blue with indigo and then use one of several yellow solutions made from various other plants. The result looked fine at first, but the yellow was usually fugitive and soon faded out leaving only the blue [16].

Green *lǜ* 綠 (etymologically “yellowish-green”) is one of the colors that life takes on and is the emblem of the spring. “Green clouds” *lǜyún* 綠雲 is an epithet for the hair of a beautiful girl.

Green was a color especially prized by Manchu women, who may have seen it as a symbol of growing things, hence expressing fertility; it was forbidden for men, except for the court eunuchs [16].

But green can also be a negative symbol. A man wearing a green turban (*lǜmàozi* 綠帽子, or *lǜtóujīn* 綠頭巾) is a cuckold and the “green mansion” (*qīnglóu* 青樓) is a brothel. In the last decade, the term “green tea bitch” (*lǜchábiāo* 綠茶婊) – formed by a metaphor for the innocent and charming appearance of a girl with long silky black hair and the formant that refers to her actual calculating and manipulative behavior – has become current in China.

The Chinese attach great importance to combinations of the two complementary colors, red and green. Painting may still be referred to as *dānqīng* 丹青 (“cinnabar-red” + “grue”) [3]. The expression “blue hills and green streams” (*qīng shān lǜ shuǐ* 青山綠水) that is a metaphor for charming natural landscape stems from the homophonous blue-and-green (*qīnglǜ* 青綠) landscape painting style. Azurite mineral blue and malachite mineral green were the principal colors used for portraying scenic elements [25]. The blue-and-green manner appears to have been specially chosen to illustrate themes of Daoist paradise, mountain retreats and imperial parks, eremitism, immortality, and alchemy [4].

To ensure a long life the Chinese have recourse to certain complicated charms, one of those is burial clothes, known as “longevity garments” (*shòuyī* 壽衣). Among these clothes, there is one robe in particular on which special pains have been lavished to imbue it with this priceless quality. It is a long silken gown of the deepest blue color, with the word “longevity” (*shòu* 壽) embroidered all over it in thread of gold. To present an aged parent with one of these costly and splendid mantles is esteemed by the Chinese an act of filial piety and a delicate mark of attention [26].

Purple

The color *zǐ* 紫 purple became sacral in Daoism: Daoist texts were referred to as “purple books” (*zǐ shū* 紫書), Laozi, according to the legend, was once accompanied by purple clouds when traveling. “Purple palace” (*zǐ gōng* 紫宮) is the term that indicates the celestial home of the supreme deity, which is believed to be the around the North Star. The Imperial Palace complex in Beijing (built in

1402–1420) is aligned with the North Star and is aimed to be the terrestrial projection of the celestial home of the supreme deity. This is the reason why its complete name is “Purple Forbidden City” (*Zǐjìnchéng* 紫禁城) [20].

During the Ming dynasty (1368–1644), purple was regarded as one of the imperial colors (together with reddish-black *xuán* 玄 and yellow *huáng* 黃) and was debarred from wearing amongst officials and common people [12]. The “purple paths” (*zǐ mò* 紫陌) are those that lead to the imperial capital, imperial edicts were referred to as “purple written admonitions” (*zǐ gào* 紫誥).

“Purple Maiden” (*Zǐgū* 紫姑) is the goddess of latrines in Chinese folk mythology. The placenta is known as the “carriage of the purple river” (*zǐhéché* 紫河車). On stage, heroes with purple faces can be recognized as loyal, imperturbable officials [3]. An extremely popular person can be described as “bright red and very purple” (*dàhóng dàzǐ* 大紅大紫).

Originally a textile term, *zǐ* 紫 covered the shades from purplish-red to purple in traditional China [22]. Nowadays, it mainly denotes violet.

Pink

Reddish-pink was originally designated by the term *hóng* 紅 and included in the RED category (see section “Red”). Since around the eleventh century, the literal meaning of the two Chinese characters that refer to the color pink is “powdered-red” (*fěnhóng* 粉紅) and “make-up red.” Characters meaning “peach-red” (*táohóng* 桃紅) are also sometimes used for “pink.” The color symbolizes female flightiness; it is also the color of fornication [3].

“Peach-blossom color” (*táosè* 桃色) figuratively means “illicit love” and is the result of loan translation from English *pink* which has the meaning “indecent, vulgar, romantic, or mildly ‘blue,’” as in *táosè xīnwén* 桃色新聞 (“pink news”) “scandalous news involving love affairs” [14].

The relatively recent term “Little Pink” (*xiǎofěnhóng* 小粉紅) that now broadly refers to young Chinese aggressive cyber-nationalists (of any gender) originally emerged on the popular women’s literature website Jinjiang Literature

City (Jinjiāng wénxué chéng 晉江文學城). Not only this online community had pink (*fěnhóng* 粉紅) as its background color, but also its content strategies deployed during the cross-strait memes war revealed that nationalistic discourses did feature an element of soft seduction, which may have been interpreted as “feminine” [27].

Conclusion

The human eye can distinguish myriads of hues, but the color vocabulary in a language is always limited. This happens because a color term is always an abstraction, and abstractions and generalizations do not remain intact at all stages, they *themselves* are the product of social and cultural development. Color in a language inspires associations and connotations that are often very subtle and determined by deep “culture memory.”

Cross-References

- [Ancient Color Terminology](#)
- [Color Symbolism](#)
- [Translation of Color Terms](#)

References

1. Lai, G.: Colors and color symbolism in early Chinese ritual art: red and black and the formation of the five colors system. In: Dusenbury, M.M. (ed.) *Color in Ancient and Medieval East Asia*, pp. 25–43. Spencer Museum of Art, University of Kansas (2015)
2. Wang, T.: Color terms in Shang oracle bone inscriptions. *BSOAS*. **59**(1), 63–101 (1996)
3. Eberhard, W.: *A Dictionary of Chinese Symbols: Hidden Symbols in Chinese Life and Thought*. Trans. G.L. Campbell. SMC Publishing Inc., Taipei (1994)
4. McNair, A.: On the meaning of the “blue-and-green manner” in Chinese landscape painting. In: Weidner, M. (ed.) *Perspectives on the Heritage of the Brush*, pp. 65–77. Spencer Museum of Art, Lawrence (1997)
5. Blänsdorf, C., Horn, F.: *Encounter in Colour: Polychromy of the Qin Terracotta Army*. Museum für Abgüsse Klassischer Bildwerke, München (2009)
6. Vainker, S.: *Chinese Silk: A Cultural History*. Rutgers University Press, New Brunswick (2004)
7. Biggam, C.P.: *The Semantics of Colour: A Historical Approach*. Cambridge University Press, Cambridge (2012)
8. Bogushevskaya, V.: Ancient Chinese ‘five colours’ theory: what does its semantic analysis reveal? In: Goldman, R.B. (ed.) *Essays in Global Color History: Interpreting the Ancient Spectrum*, pp. 225–244. Gorgias Press, Piscataway (2016)
9. Baxter, W.H.I.I.: A look at the history of Chinese color terminology. *J. Chin. Lang. Teach. Assoc.* **XIX**(2), 1–26 (1983)
10. Chêng, T.: The t’u-lu colour-container of the Shang-Chou period. *BMFEA*. **37**, 239–250 (1965)
11. Wenren, J.: *Ancient Chinese Encyclopaedia of Technology: Translation and Annotation of the Kaogongji (the Artificers’ Record)*. Routledge, New York (2013)
12. Han, J.: *The historical and chemical investigation of dyes in high status Chinese costume and textiles of the Ming and Qing dynasties (1368–1911)*. PhD thesis, University of Glasgow (2016)
13. Dusenbury, M.M.: Introduction. In: Dusenbury, M.M. (ed.) *Color in Ancient and Medieval East Asia*, pp. 11–15. Spencer Museum of Art, University of Kansas (2015)
14. Xu, W.: *A Study of Chinese Color Terminology*. Lincom, Muenchen (2007)
15. Thieme, C.: Paint layers and pigments on the Terracotta Army: a comparison with other cultures of antiquity. In: Wu, Y., Zhang, T., Petzet, M., Emmerling, E., Blänsdorf, C. (eds.) *The Polychromy of Antique Sculptures and the Terracotta Army of the First Chinese Emperor*, pp. 52–58. ICOMOS, München (2001)
16. Cammann, S.: Costume in China, 1644 to 1912. *Philadelphia Museum Art Bull.* **75**(326), 3–19 (1979)
17. Kieschnick, J.: The symbolism of the monk’s robe in China. *Asia Major*. **12**(1), 9–32 (1999)
18. Liú, Y.: *Yǔyán de sècǎi měi 語言的色彩美 [The Beauty of Colours in a Language]*. Ānhuī jiàoyù, Héféi (1990)
19. Bogushevskaya, V.: Formal theory-driven, psycholinguistic data and corpus-driven study confirms the absence of a basic colour term for ORANGE in Modern Standard Mandarin and elaborates the syntactico-semantic ‘distributional potential’ criterion for basicness. *L’Analisi Linguistica e Letteraria*. **2**, 61–72 (2020)
20. Wú, D.: *Sècǎi yǔ zhōngguó rén de shēnghuó 色彩與中國人的生活 [Colours in Chinese People’s Life]*. Tuánjié, Běijīng (2000)
21. Bonds, A.B.: *Beijing Opera Costumes: The Visual Communication of Character and Culture*. University of Hawai’i Press, Honolulu (2008)
22. Boguševskaja, V.: *Semantika cvetonaimenovaniy v kitajskom jazyke (universal’noe i nacional’noe) [The semantics of colour terms in Chinese: universal and areal characteristics]*. PhD dissertation, Moscow State University (2008)
23. Han, J., Quye, A.: *Dyes and dyeing in the Ming and Qing dynasties in China: preliminary evidence based*

- on primary sources of documented recipes. *Text. Hist.* **49**(1), 44–70 (2018)
24. Bogushevskaya, V.: GRUE in Chinese: on the original meaning and evolution of qīng (青). In: Bogushevskaya, V., Colla, E. (eds.) *Thinking Colours: Perception, Translation and Representation*, pp. 26–44. Cambridge Scholars Publishing, Newcastle upon Tyne (2015)
 25. Cheng, M., Tang, W.H., Choy, E.: *Essential terms of Chinese painting*. City University of Hong Kong Press, Hong Kong (2017)
 26. Frazer, J.G.: *The Golden Bough: A Study of Magic and Religion*. Cosimo Classics, New York (2009)
 27. Fang, K., Repnikova, M.: Demystifying “little pink”: the creation and evolution of a gendered label for nationalistic activists in China. *New Media Soc.* **20**(6), 2162–2185 (2018). <https://doi.org/10.1177/1461444817731923>

Chromatic Contrast

► Color Contrast

Chromatic Contrast Sensitivity

Christoph Witzel¹ and Karl Gegenfurtner²

¹School of Psychology, University of Southampton, Southampton, UK

²Department of Psychology, Justus-Liebig University, Giessen, Germany

Synonyms

Color Contrast Sensitivity (rarely)

Definition

Chromatic contrast sensitivity is the ability to see chromatic contrast. Chromatic contrast refers to the occurrence of differences in *chromaticity* (saturated, hueful color) in a visual percept (scene, image, and stimulus). It may consist in differences across space (*spatial chromatic contrast*) or in changes of chromaticity across time

(*temporal chromatic contrast*). The term chromatic contrast is used in opposition to *achromatic contrast*, where differences only occur in *luminance* (greyscale). Chromatic contrast sensitivity depends on the spatial and temporal frequency of contrast (low-pass) in a way that differs from achromatic contrast sensitivity (band-pass).

Conceptual Clarifications

Almost every phenomenon in color vision involves contrasts between colors. Colors are not perceived absolutely, but relative to other colors, and contrasts between colors affect the appearance of the single colors (for review, see [1]). Nevertheless, the term *chromatic contrast* has been associated with certain aspects in the perception of color differences rather than others (for review see [2]). Conceptual connotations and distinctions in different contexts are clarified below.

Achromatic Contrast, Isochromaticity, and Isoluminance

Since human color vision involves an achromatic luminance dimension and two chromatic dimensions, visual contrasts may occur in terms of luminance or chromaticity. The term chromatic contrast must be understood as the complement of **achromatic contrast** [2]. Achromatic contrast refers to differences in luminance, which are perceived as lightness differences. Spatial achromatic contrasts may be illustrated by grey-scale images, such as black-and-white photos. Unlike achromatic contrast, chromatic contrast involves differences in chromaticity, which are differences along one or both of the chromatic dimensions as opposed to the achromatic dimension.

A chromatic stimulus may consist of purely achromatic contrasts, for example when the paper of a black-and-white photo yellowed over time, or when a grey-scale image was printed on color paper. Since such images only contain lightness differences they are void of chromatic contrast [3]. Such an image (or stimulus) is called **isochromatic** because it is constant (iso ~ equal) in chromaticity [4]. In contrast, an **isoluminant** image is constant in luminance and only varies in

chromaticities, implying that it has chromatic contrast while its achromatic contrast is zero.

Color Contrast

Because chromatic and achromatic differences are both part of human color perception, “color contrast” may appear to conceptually include both, achromatic and chromatic contrast. Nevertheless, color contrast is often used interchangeably with chromatic contrast in the literature depending on different research domains. The term **chromatic contrast** is mostly (but not exclusively) used in research on *contrast sensitivity* and its relationship to spatial and temporal frequency [2]. For this reason, the term chromatic contrast focuses on the distribution of color differences across space and time, and the ability of the observer to detect these differences (contrast sensitivity). On the other hand, the term **color contrast** is predominantly used in the context of research on *chromatic induction* that is the effect of color contrasts on *color appearance* (the subjective impression of color). In that context, *lightness contrast* is the greyscale complement of color contrast.

Color Edges, Spatial and Temporal Contrast

Chromatic (and achromatic) contrast may occur through variations across space or across time [2]. On the one hand, **spatial chromatic contrast** consists in the simultaneous occurrence of differences in chromaticity at different locations in space. On the other hand, **temporal chromatic contrast** refers to changes in chromaticity over time. In both cases the strength of the contrast is defined as the size of the difference in chromaticity.

The perception of contrast strongly depends on the transition between the two chromaticities. Transitions can be rather gradual or abrupt. For example, the chromatic difference between a raspberry and the surrounding leaves provides an abrupt transition between red and green, while the transitions between the colors of a rainbow are gradual. The temporal transition between the colors of a traffic light is abrupt, while the change of the chromaticity of sunlight during the day is very slow and gradual. Such transitions may be assessed by the spatial and temporal frequency of the color transition, respectively. The ability to

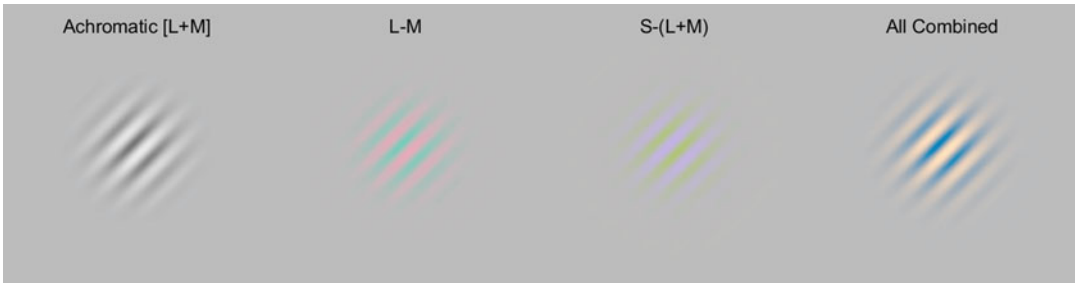
perceive chromatic contrast depends on these spatial [3] and temporal frequencies [5].

Chromatic edges (or *chromatic edge contrast*) are a particular kind of spatial chromatic contrast. Edge contrasts are related to the visual environment (*scenes*) because they are delimiting objects or other fundamental features of a scene such as shadows or highlights. As a result, while the term chromatic contrast focuses on the sensitivity to color contrasts in general, the notion of chromatic edges highlights the relationship between the beholder and their visual environment, and is rather used in the context of *scene statistics* (for review, see [1], e.g. [6, 7]).

Detection, Discrimination, and Contrast Sensitivity

Chromatic contrast sensitivity refers to the ability to see chromatic contrast. The term contrast sensitivity accounts for the dependence of sensitivity on the temporal and spatial distribution of contrasts. It is determined as the minimum contrast that an observer can see, given a specific temporal and spatial frequency. For this purpose, chromatic contrast sensitivity is measured with periodical changes in chromaticity, mostly involving two opponent chromaticities (see Fig. 1 and Method).

Chromatic sensitivity, color detection, and color discrimination involve chromatic contrast sensitivity; but they typically refer to differences in chromaticity independent of spatial or temporal frequency and are measured with single presentations of uniformly colored areas, typically colored disks (e.g. [8, 9]). **Chromatic sensitivity** is the general ability to see chromatic colors and is measured through color detection. **Color detection** consists in the discrimination of one color (*test*) from the neutral (i.e., *adapting*) background. Chromatic sensitivity is determined as the minimum difference in chromaticity between test color and background that can be detected. **Color discrimination** consists in the discrimination between two different colors (*test* and *comparison*) shown on the adapting background. The ability to discriminate colors (*sensitivity to color differences*) is determined through the minimum difference between the colors that can be discriminated on a given background. Color detection may be conceived as



Chromatic Contrast Sensitivity, Fig. 1 Typical Stimuli to Measure Spatial Contrast Sensitivity. From left to right: Gabor patches with achromatic contrast, L-M, S-(L + M), and all three contrasts combined. Note that the chromatic contrasts are more visible when looking from a closer

distance. This illustrates the low-pass frequency of the spatial chromatic contrast sensitivity function: Moving the observer toward the images makes the image larger and hence decreases spatial frequency

a special case of color discrimination, in which the test color is identical with the background.

Color detection and discrimination may be understood as special cases of chromatic contrast sensitivity. Chromatic contrast sensitivity converges to a maximum at low spatial and temporal frequencies and is zero above an upper boundary (*chromatic acuity*) of spatial frequencies (see Chromatic contrast sensitivity functions). By using sufficiently large color patches (> 1 degree visual angle) and presenting them for a sufficiently long time (> 200 ms) contrast sensitivity will be at maximum, independent of additional high spatial frequency (edges of stimuli) and high temporal frequency components (abrupt stimulus onset). In this way, color detection and discrimination performance reflects maximal chromatic contrast sensitivity in terms of spatial and temporal frequency, and does not require further specification of spatial and temporal frequency components in the stimulus (e.g. [8, 9]).

Spatial Resolution and Chromatic Acuity

Spatial resolution refers to the maximal spatial frequency that can still be seen, such as the minimum size of letters that can still be identified. Because of the tight relationship between spatial frequency and contrast sensitivity, contrast is also the complement of spatial resolution and acuity. For example, the smallest letters that can still be read when printed in black on a white paper, may not be identified when printed in light grey instead of black, because they have lower contrast than

the black letters. Hence, the spatial resolution depends on the contrast. This is also true for the spatial resolution of color vision. It can only be determined relative to a given contrast: The thinnest color line that may be detected with a high chromatic contrast, will be invisible with a lower contrast.

Visual acuity is a particular case of spatial resolution, which corresponds to the highest spatial frequency that can be detected at maximum contrast [2]. It can only be determined relative to a given contrast, such as between black letters and white background (cf. Snellen chart to measure visual acuity). Analogically, chromatic acuity corresponds to the highest spatial frequency that is still visible when presented with maximum chromatic contrast (e.g. [3]).

Method

Stimuli to measure chromatic contrast sensitivity vary periodically in chromaticity but are iso-luminant to separate achromatic from chromatic contrast (Fig. 1). To quantify chromatic contrast, differences between chromaticities need to be calculated.

Contrast Indices

Contrast indices calculate differences in chromaticities relative to a reference intensity, which is supposed to correspond to the adaptation of the

observer. Mainly three indices of contrast have been used.

The **Weber contrast** is calculated by the difference in intensity between two colors divided by the intensity of the background. This index is sensible if the background corresponds to the observer's adaptation color and the perception of the difference depends on the intensity of this background (e.g. [8, 9]).

Alternatively, the **Michelson contrast** divides the difference between intensities by the sum of their intensities. This approach is sensible when it may be assumed that observers are adapted to the intensities of the two colors, rather than to the background. Both indices are equivalent when the two colors vary symmetrically around the background. These two indices are useful for specifying contrast in sine-wave gratings, for example when measuring contrast sensitivity (e.g. [3, 10]).

Finally, the *Root Mean square* or **RMS contrast** consists in the standard deviation of the chromatic differences across a scene. In order to relativize these differences to the overall mean, the values are standardized before the calculation of the RMS index. This index may not only be used for contrast sensitivity measurements, but also for contrast distributions in scenes. However, still other indices to assess contrast are possible [2].

Color Differences

Color does not directly map to a physical measure (for review, see [1]). As with color in general, the quantification of chromatic contrast is relative to perception. Consequently, metrics of chromatic contrast are relative to the dimensions used to quantify chromaticity ("intensity of color"). For the measurement of contrast sensitivity, two ways to quantify chromatic contrast have been used.

First, contrast may be determined by the **physical** or *instrument contrast* [11]. The instrument contrast is the relative intensity of two chromatic lights (*component colors*), such as two monitor primaries or two monochromatic lights (e.g. [3]). The contrast between the two unmixed component colors corresponds to 100%, and they are mixed to produce intermediate levels of contrast (cf. Michelson contrast). The physical intensity of

each component color is typically measured as proportional luminance. However, the component colors, the dimension along which the difference is measured, and the scaling of the differences are arbitrary and depend on the device used to produce the colors.

Second, **cone-contrast** is used as a perceptual measure of contrast (for review, see [2], e.g. [12]). Cone contrasts are calculated by the difference in absorption (or excitation) of each of the three photoreceptors (L-, M- and S-cones) relative to the state of adaptation of that photoreceptor. Weber or Michelson contrast may be used for this purpose. Alternatively, colors may be directly represented in cone-contrast space, in which each of the three dimensions (ΔLMS) reflects the cone-contrast for one type of cone.

The most widespread approach is to determine contrast sensitivities relative to the **cone-opponent signal** of the postreceptoral, second-stage mechanisms (for review, see [2], e.g. [13, 14]). Chromatic contrast is then quantified in terms of the L-M and S-(L + M) cone contrast, such as in Fig. 1 (for review, see [15]). An advantage of this approach is that cone-opponent chromatic contrasts are orthogonal at isoluminance. The RMS contrast may be used to combine the contrasts across cone-types or opponent dimensions, such as in the rightmost panel of Fig. 1 (e.g. [12, 16]).

Note, that perceptual quantifications depend on the perceptual processes that are modeled by the difference metric and on the knowledge about these processes (for review, see [1]). Cone-contrasts relate chromatic contrast sensitivity to the low-level, cone-opponent processes of color vision. Hence, comparisons in contrast across different chromaticities are relative to these mechanisms. Higher-level, cortical mechanisms might involve different kinds or reweighed contrasts, and hence quantifications of contrast may differ for high-level mechanisms. In addition, studies may differ in how they define maximum cone-contrast, which affects the relative scaling of the cone-opponent axes.

Spatial Frequency

Gratings that alternate periodically between contrasting colors are used to control spatial

frequencies (Fig. 1). If stimuli consisted of simple uniform bars, sharp edges would occur between the single bars of the grating and between the grating and the background. Such sharp edges contain high spatial frequency components, which would be confounded with the low spatial frequency of the bars.

To avoid such edges between the bars, gratings are sinusoidally modulated [3, 10]. To further avoid sharp edges between grating and background contrasts are gradually reduced toward the edges through the application of a filter. As a result, the grating smoothly melts into the background and the strongest stimulation (i.e., contrast) is at the center of the grating. The molding of the overall grating through a filter is called an **envelope**, and in most cases this filter consists in a Gaussian function (*Gaussian envelope*). A sinusoidally modulated grating with an envelope is called a *Gabor patch* [17, 18]. Figure 1 provides examples of Gabor patches. Code for creating Gabor patches may be found in the PsychDemos of the Psychtoolbox [19].

While contrast may be controlled as the amplitude of the sinusoids, spatial frequency corresponds to the frequency of cycles. Since the perceived spatial frequency depends on the retinal image, it also depends on the distance of the eye to the grating (cf. Fig. 1). For this reason, spatial frequencies are quantified as *cycles per degree* (cpd) of visual angle. The visual angle accounts for the distance. At 0.5 cpd, one cycle covers 2 degrees, which is approximately the size of the fovea. Two degree visual angle roughly correspond to a width of 2cm at a distance of 57cm, or the size of a thumb at arm's length.

Note that there is a trade-off between controlling the number of cycles and the size of stimuli [3, 18]. Stimuli should be large enough to show a sufficient number of cycles for all frequencies (e.g. 2.5 cycles, cf. [2, 18]). For gratings above 3–4 cpd chromatic aberration produces luminance artifacts that influence detection thresholds. This is mainly the case for blue-yellow gratings, since the wavelength compositions for red and green chromaticities do not differ strongly, resulting in a similar refraction [3, 4].

Classical studies used absent-present judgments [3] or adjustment tasks [10] to measure chromatic contrast sensitivity, often with experienced observers. Instead, alternative forced choice tasks avoid relying on observer criteria. For example, gratings may be presented at different time points (e.g. [14]), at different locations of the display (e.g. [16]), or with different orientations (e.g. [13]), and observers indicate the time point, location or, respectively, the orientation of the grating through a forced-choice response.

Temporal Frequency

High temporal frequencies correspond to fast flicker between contrasting chromaticities, low frequencies to slow, and gradual transitions. Temporal frequency is quantified as cycles per second that is *Hertz* (Hz).

Temporal contrast sensitivity is measured through **heterochromatic flicker** [5]. In this setup, chromaticities periodically change, and temporal transitions are sinusoidally modulated to control specific temporal frequencies. At high frequencies, chromaticities are seen as an unchanging mixture of contrasting chromaticities (*flicker-fusion*). At extremely low frequencies, the change in chromaticity is not visible, either. For example, the slow change in chromaticity of daylight remains usually unnoticed in everyday life. Note that achromatic artifacts occur in L-M flicker due to latency differences between L- and M-cones [2].

Spatiotemporal contrast sensitivity, i.e., contrast sensitivity as a function of spatial and temporal frequency, may be measured by oscillating between gratings that are phase-shifted by 90 degree and hence show spatially inverse (green-red-green vs. red-green-red) contrasts [10].

Chromatic Contrast Sensitivity Functions

The relationship between contrast sensitivity and spatial and temporal frequency is captured through chromatic contrast sensitivity functions (cCSF). In CSFs, contrast sensitivity is represented along the y-axis as a function of spatial frequency (*spatial contrast sensitivity*

function, sCSF), or temporal frequency (*temporal contrast sensitivity function*, tCSF). All axes are usually scaled logarithmically.

Spatial Contrast Sensitivity

Chromatic spatial contrast sensitivity functions are **low-pass** [3, 4, 13]. This means that contrast sensitivity declines with high spatial frequencies (“narrower lines”), but barely changes when spatial frequency approaches zero, i.e., when the transition between contrasting colors is completely gradual. The chromatic sensitivity function reaches its maximum at about 0.5 cpd, i.e., when one cycle covers the whole fovea (see above). This low-pass pattern is largely independent of absolute luminance [16].

The low-pass pattern contrasts the one found for achromatic sCSFs. Achromatic sCSFs are band-pass with a peak at about 3–5 cpd and decreasing sensitivity toward both, higher and lower spatial frequencies. At the same time, acuity is much higher for achromatic (cut-off at 40–60 cpd) than for L-M (about 12 cpd) and S-(L + M) contrasts (10 cpd). In fact, due to chromatic aberration the effective resolution is only 3–4 cpd for the S-(L + M) contrast [2]. Moreover, increasing eccentricity from the fovea toward the periphery has a stronger attenuating effect on L-M contrast sensitivity than on achromatic contrast sensitivity, while S-(L + M) contrast sensitivity declines with eccentricity in a similar way as achromatic contrast sensitivity [2, 14].

These properties of spatial contrast sensitivity imply that “very thin” stripes are still visible when shown in black and white, but “disappear” when shown as isoluminant colors [3, 16]. This has been used, for example, in image compression (e.g., TV broadcast) where high-spatial frequency information is saved in achromatic contrast only and chromatic high-spatial frequency information is discarded [20]. However, chromatic contrast sensitivity is overall higher than achromatic contrast sensitivity when expressed as cone-contrasts instead of physical contrasts [12].

Temporal Contrast Sensitivity Function

Temporal frequency modulates chromatic contrast sensitivity in a similar way as spatial frequency [2, 5]. It is mainly **low-pass**, with only a slight attenuation in sensitivity below about 1 cpd.

It is highest when colors change 1–3 times per second (1–3 Hz), and temporal resolution is at about 30–40 Hz. In contrast, achromatic contrast sensitivity peaks between 5 and 20 Hz, and has a higher cut-off value for temporal resolution (~50 Hz). Since both, spatial and temporal cCSFs are low-pass, the spatiotemporal cCSF is also low-pass [10].

Due to the lower temporal resolution for differences in chromaticity, heterochromatic flicker can be used to exactly determine isoluminance for an individual observer (**heterochromatic flicker photometry**). At a temporal frequency of 15–20 Hz chromaticities of contrasting colors fuse, and the only flicker left is achromatic. The brightness of one of the contrasting colors is adjusted until there is no flicker, which implies that the luminance of the two colors is equal [21].

Development

As with adults, the spatial (L-M) CSF is low-pass in infancy (8–30 weeks), and develops through a steady increase in overall sensitivity and in spatial resolution [22]. In contrast, temporal (L-M) CSFs are band-pass in 3 month old infants, and develop an adult-like low-pass profile only at the age of 4 months [11]. Across the life-span, chromatic contrast sensitivity increases steadily until adolescence, and decreases thereafter [23].

Eye Movements

While sensitivity for luminance contrast decreases during *smooth pursuit*, sensitivity for isoluminant L-M and S-(L + M) contrast increases (10–16%), starting about 50 ms before eye movement [24]. An increase in contrast sensitivity has also been found during *optokinetic nystagmus*, but not *vestibulo-ocular reflex* ([25], for review, see [26]).

Theories on Chromatic Contrast Sensitivity

Physiological Basis

The optics of the eye blur the retinal image and explain a major part of sensitivity loss for high spatial frequencies for both, chromatic and achromatic contrasts. Blurring through chromatic aberration further reduces S-(L + M) contrasts of high

spatial frequency in the retinal image. In addition, the distribution of cones across the retina (*cone mosaic*) modulates contrast sensitivity. The scarcity of S-cones in the retina further attenuates the spatial resolution of S-(L + M) contrast sensitivity [4].

Chromatic contrasts are processed by two independent **pathways** that go from the *retinal ganglion cells* via the *lateral geniculate nucleus* (LGN) to the *visual cortex*. The *parvocellular* pathway processes L-M, the *koniocellular* pathway S-(L + M) contrasts. The decrease in contrast sensitivity with eccentricity agrees with the distribution of parvocellular and koniocellular receptive fields across the retina [2, 14]. Moreover, the increase of chromatic contrast sensitivity during pursuit and optokinetic nystagmus indicates a boost of the parvo- and koniocellular system [24, 26].

Functional Role in Ecology

Chromatic contrasts are statistically independent from luminance contrast in natural scenes, and hence provide an additional source of information for object identification apart from luminance contrast [6, 7]. Furthermore, spatial L-M contrast sensitivity is optimal for the detection of red and yellow fruits at reaching distance (for review, see [1, 27]).

Cross-References

- [Color Combination](#)
- [Color Contrast](#)
- [Color Processing, Cortical](#)
- [Color Scene Statistics, Chromatic Scene Statistics](#)
- [Complementary Colors](#)
- [Cortical](#)
- [Photoreceptors, Color Vision](#)

References

1. Witzel, C., Gegenfurtner, K.R.: Color perception: objects, Constancy, and categories. *Annu. Rev. Vis. Sci.* **4**(1), 475–499 (2018)
2. Diez-Ajenjo, M.A., Capilla, P.: Spatio-temporal contrast sensitivity in the cardinal directions of the colour space. *A Review. J. Optom.* **3**(1), 2–19 (2010)
3. Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol.* **359**, 381–400 (1985)
4. Williams, D., Sekiguchi, N., Brainard, D.: Color, contrast sensitivity, and the cone mosaic. *Proc. Natl. Acad. Sci. U. S. A.* **90**(21), 9770–9777 (1993)
5. Kelly, D.H., van Norren, D.: Two-band model of heterochromatic flicker. *J. Opt. Soc. Am.* **67**(8), 1081–1091 (1977)
6. Hansen, T., Gegenfurtner, K.R.: Independence of color and luminance edges in natural scenes. *Vis. Neurosci.* **26**(1), 35–49 (2009)
7. Hansen, T., Gegenfurtner, K.R.: Color contributes to object-contour perception in natural scenes. *J. Vis.* **17**(3), 14–14 (2017)
8. Krauskopf, J., Gegenfurtner, K.R.: Color discrimination and adaptation. *Vis. Res.* **32**(11), 2165–2175 (1992)
9. Witzel, C., Maule, J., Franklin, A.: Red, yellow, green, and blue are not particularly colorful. *J. Vis.* **19**(14), 27–27 (2019)
10. Kelly, D.H.: Spatiotemporal variation of chromatic and achromatic contrast thresholds. *J. Opt. Soc. Am.* **73**(6), 742–750 (1983)
11. Dobkins, K.R., Anderson, C.M., Lia, B.: Infant temporal contrast sensitivity functions (tCSFs) mature earlier for luminance than for chromatic stimuli: evidence for precocious magnocellular development? *Vis. Res.* **39**(19), 3223–3239 (1999)
12. Chaparro, A., et al.: Colour is what the eye sees best. *Nature.* **361**(6410), 348–350 (1993)
13. Webster, M.A., De Valois, K.K., Switkes, E.: Orientation and spatial-frequency discrimination for luminance and chromatic gratings. *J. Opt. Soc. Am. A.* **7**(6), 1034–1049 (1990)
14. Mullen, K.T., Kingdom, F.A.: Differential distributions of red-green and blue-yellow cone opponency across the visual field. *Vis. Neurosci.* **19**(1), 109–118 (2002)
15. Gegenfurtner, K.R., Kiper, D.C.: Color vision. *Annu. Rev. Neurosci.* **26**(1), 181–206 (2003)
16. Wuerger, S., et al.: Spatio-chromatic contrast sensitivity under mesopic and photopic light levels. *J. Vis.* **20**(4), 23–23 (2020)
17. Gabor, D.: Theory of communication. Part 1: the analysis of information. *J. Inst. Elect. Eng. Part III.* **93**(26), 429–441 (1946)
18. Caelli, T., Moraglia, G.: On the detection of Gabor signals and discrimination of Gabor textures. *Vis. Res.* **25**(5), 671–684 (1985)
19. Kleiner, M., Brainard, D., Pelli, D.: What's new in psychtoolbox-3? *Perception.* **36** (2007) ECVF Abstract Supplement
20. Watson, A.B.: Perceptual-components architecture for digital video. *J. Opt. Soc. Am. A Opt. Image Sci.* **7**(10), 1943–1954 (1990)

21. Pokorny, J., Smith, V.C., Lutze, M.: Heterochromatic modulation photometry. *J. Opt. Soc. Am. A.* **6**(10), 1618–1623 (1989)
22. Teller, D.Y.: Spatial and temporal aspects of infant color vision. *Vis. Res.* **38**(21), 3275–3282 (1998)
23. Knoblauch, K., Vital-Durand, F., Barbur, J.L.: Variation of chromatic sensitivity across the life span. *Vis. Res.* **41**(1), 23–36 (2001)
24. Schütz, A.C., et al.: Improved visual sensitivity during smooth pursuit eye movements. *Nat. Neurosci.* **11**(10), 1211–1216 (2008)
25. Schütz, A.C., Braun, D.I., Gegenfurtner, K.R.: Chromatic contrast sensitivity during optokinetic nystagmus, visually enhanced vestibulo-ocular reflex, and smooth pursuit eye movements. *J. Neurophysiol.* **101**(5), 2317–2327 (2009)
26. Schütz, A.C., Braun, D.I., Gegenfurtner, K.R.: Eye movements and perception: a selective review. *J. Vis.* **11**(5) (2011) Article 9
27. Parraga, C.A., Troscianko, T., Tolhurst, D.J.: Spatiochromatic properties of natural images and human vision. *Curr. Biol.* **12**(6), 483–487 (2002)

Chromatic Description

- [Color Name Applications in Computer Vision](#)

Chromatic Image Statistics

- [Color Scene Statistics, Chromatic Scene Statistics](#)

Chromatic Labeling

- [Color Name Applications in Computer Vision](#)

Chromatic Processing

- [Color Processing, Cortical](#)

Chromophore

Malgorzata Szymczyk

Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Definition

The term chromophore was introduced by Otto N. Witt in 1876, as a combination of ancient Greek terms *chromo* (color) and *phore* (bearer) [1]. In the field of color science, chromophore refers to the structural unit in an organic molecule that is responsible for color. In other words, chromophores enable organic molecules to absorb light in the visible region (400–700 nm) of the electromagnetic spectrum.

Basis for Color in Organic Compounds

Unlike most organic compounds, dyes possess color because they absorb light in the visible spectrum, have at least one chromophore, and have a conjugated system. When any one of these features is lacking from the molecular structure, color does not exist [2].

Absorption of Visible Light

When visible light strikes a colored object, certain wavelengths of that light are absorbed, and the remaining wavelengths are reflected and can be detected by the human eye. The colors associated with the absorbed and observed light are shown in Table 1. The colors absorbed and observed at a given wavelength are known as complementary colors. Combining the two colors produces white light again.

Prior to the development of synthetic colorants, color verses molecular structure in organic compounds was illustrated through conjugated olefins (see Table 2). As the length of the conjugated system increased, the absorption intensified and the maximum absorption wavelength lengthened. Here, a comparison of the colors of 1,3,5-heptatriene (colorless), Vitamin A (yellow), and

β -Carotene (orange) illustrate the importance of extended conjugation in this chromophoric system. It is also clear that many conjugated double bonds must be assembled in order to produce a deep color such as blue. As the next section will show, the development and use of chromophores more readily led to a variety of deep colors even with fewer conjugated double bonds.

Chromophore, Table 1 Wavelengths of absorbed and reflected colors

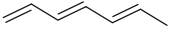
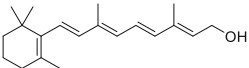
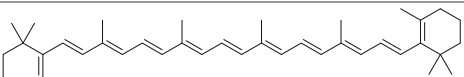
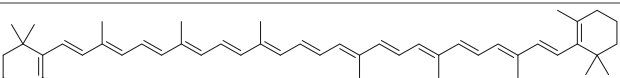
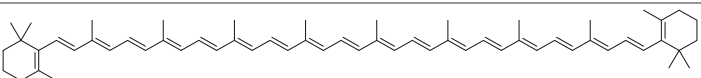
Wavelength absorbed (nm)	Color absorbed	Color reflected (observed)
400–435	Violet	Yellow–green
435–480	Blue	Yellow
480–490	Green–blue	Orange
490–500	Blue–green	Red
500–560	Green	Purple
560–580	Yellow–green	Violet
580–595	Yellow	Blue
595–605	Orange	Green–blue
605–700	Red	Blue–green

Chromophores and Conjugation in Organic Colorants

Examples of well-known chromophores in color chemistry are shown in Fig. 1, which include simple (nitro, azo, and methine) and more complex (anthraquinone and triarylmethane) systems. About 70% of synthetic dyes have the azo chromophore, with anthraquinone dyes comprising about 15% [3]. As illustrated in Fig. 2, chromophores are effective when incorporated into a conjugated system – a series of alternating single and double bonds [2, 3]. In the present example, the azo group gives color when placed between two phenyl groups but not when placed between methyl groups. In commercial dyes, the azo chromophore is produced by combining a coupler such as 2-naphthol (**1**) and a diazonium compound (**2**), as shown in Fig. 3. The formation of azo dyes such as **3** illustrates that the azo chromophore ($-\text{N}=\text{N}-$) often exists in combination with the corresponding hydrazo ($=\text{N}-\text{NH}-$) form (e.g., **4**). The two forms give overlapping absorption bands in the visible spectrum, with the hydrazo form producing the band at higher wavelength.

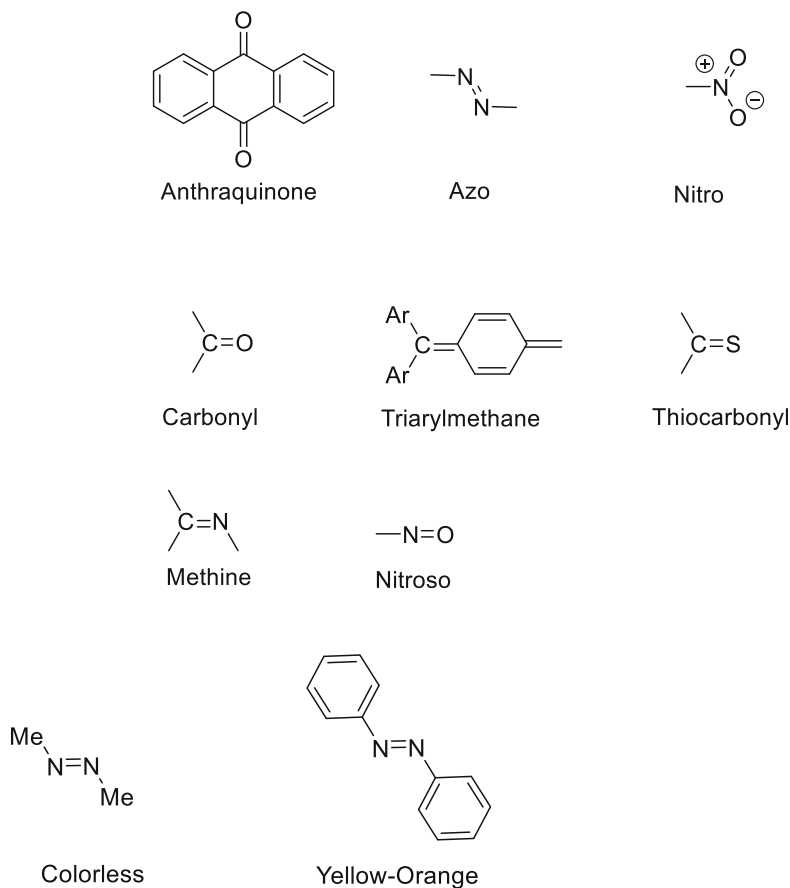
The azo chromophore affords a full scope of colors (yellow–black) and can vary in number in a given dye structure, as illustrated through textile

Chromophore, Table 2 Influence of conjugated double bonds on color in organic compounds

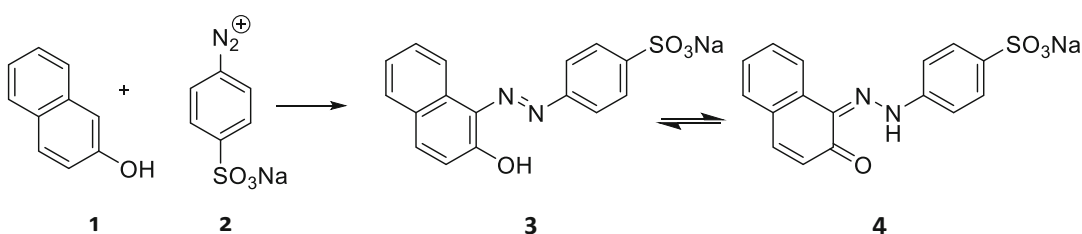
Poly(olefin) compounds	Number of conjugated –C=C– bonds	Color
 1,3,5-Heptatriene	3	Colorless
 Retinol (vitamin A)	5	Light yellow
 β -Carotene	11	Orange
 Decapreno- β -carotene-d10	15	Dark blue
 Dodecapreno- β -carotene	19	Black

Chromophore,

Fig. 1 Examples of chromophoric groups in organic colorants



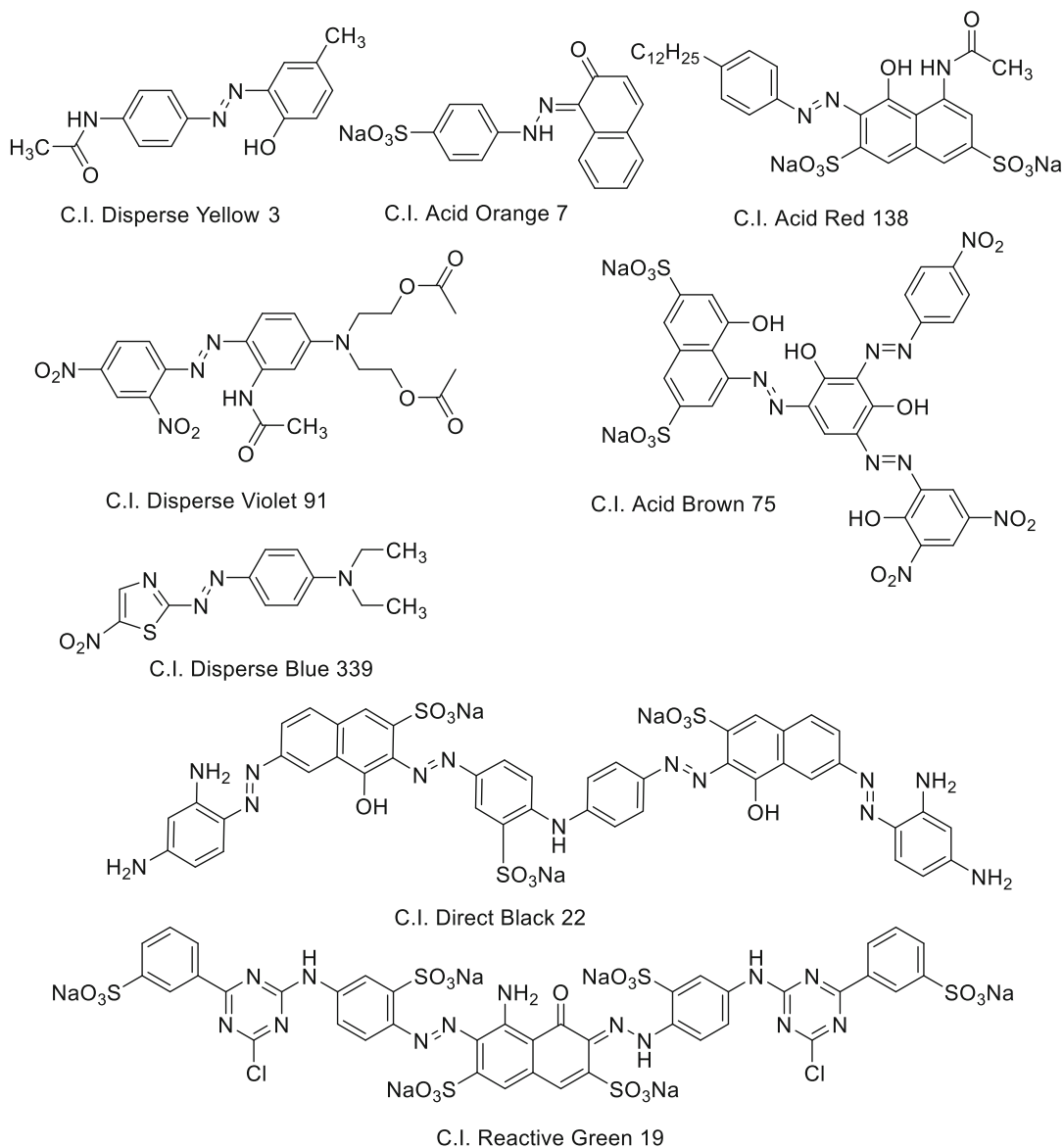
Chromophore, Fig. 2 Effects of placing an azo group between saturated and unsaturated (conjugated) systems



Chromophore, Fig. 3 Illustration of azo group formation and subsequent hydrazone tautomer

dyes C. I. Disperse Yellow 3, C. I. Reactive Green 19, C. I. Acid Brown 75, and C. I. Direct Black 22 (Fig. 4). These four examples also show that azo dyes are used in a variety of dye application classes, providing colorants having affinity for all major fiber types.

The anthraquinone chromophore gives mostly yellow–green dyes (e.g., Fig. 5), with a few examples of brown and black anthraquinones found among the vat dyes for cotton. Generally, the anthraquinone chromophore gives brighter colors and better photostability than the azo chromophore, especially among the red, blue, and green

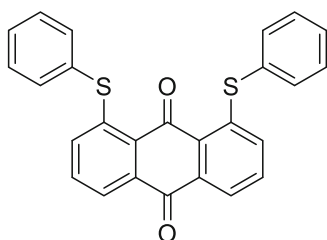


Chromophore, Fig. 4 Examples of commercial yellow–black dyes containing one to four azo groups

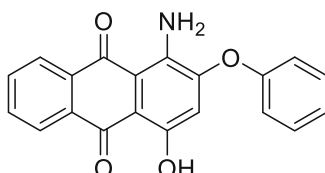
dyes. Like azo dyes, anthraquinone dyes cross many application classes, giving colorants for all major textile fiber types. Because of their brightness, anthraquinone-based products are complementary to azo dyes, which can be rather dull. They are more expensive than azo colorants, however. While atypical, dyes having an azo and an anthraquinone chromophore exist. They enable the formation of mixed colors by combining, for

example, an anthraquinone blue system with an azo yellow system (see C. I. Reactive Green 6).

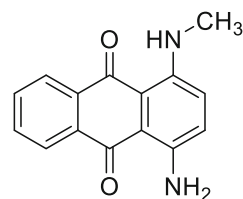
The triarylmethane chromophore gives mainly red–green dyes (e.g., Fig. 6). Triarylmethane dyes are distinguished for their bright colors and high color strength but they typically have poor photostability. The chromophoric system consists of a central carbon atom joined to three aromatic rings, generally with electron-donating groups in the



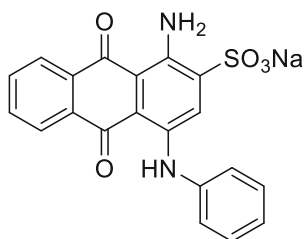
C.I. Solvent Yellow 163



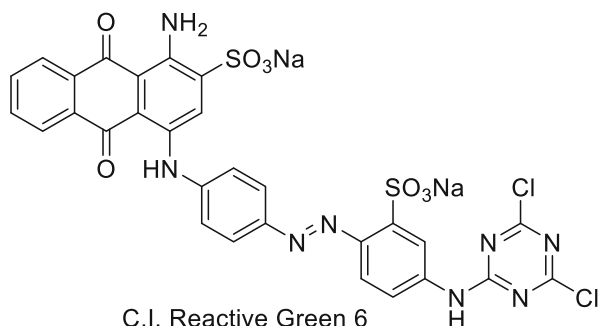
C.I. Disperse Red 60



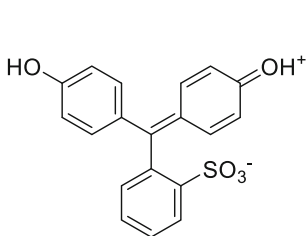
C.I. Disperse Violet 4



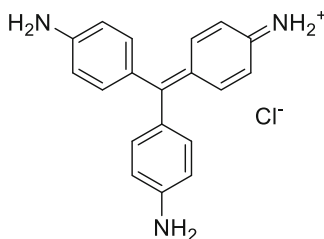
C.I. Acid Blue 25



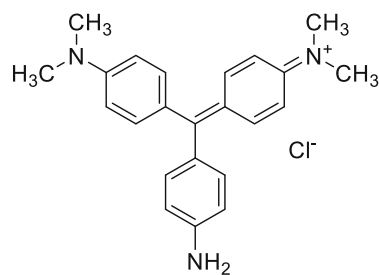
C.I. Reactive Green 6

Chromophore, Fig. 5 Examples of commercial yellow–green anthraquinone dyes

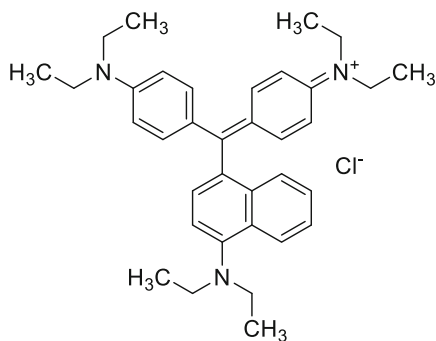
Phenol Red



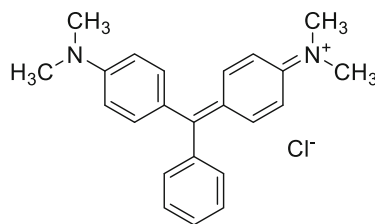
C.I. Basic Red 9



C.I. Basic Violet 1

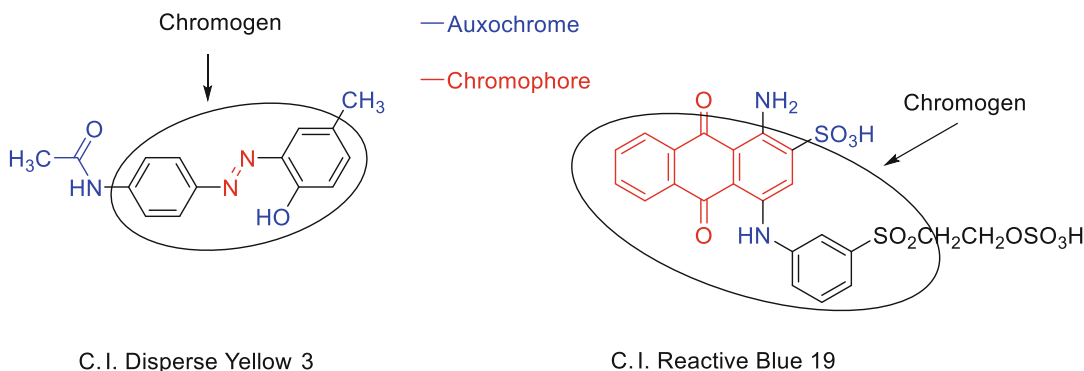


C.I. Basic Blue 7



C.I. Basic Green 4

Chromophore, Fig. 6 Examples of commercial red–green triarylmethane-based dyes



Chromophore, Fig. 7 Examples of azo and anthraquinone-based chromogens in commercial dyes

para-position. Unlike azo and anthraquinone chromophores, the triarylmethane chromophore is not used with textile coloration in view. The more common targets are inks, cosmetics, and biological stains.

The combination of chromophores and conjugated groups produces colored systems referred to as chromogens [4]. Examples of chromogens in a pair of commercial dyes are given in Fig. 7. Figure 7 also shows that commercial dyes contain groups known as auxochromes – color aiders. These groups can influence color [1, 4] and other dye properties (e.g., solubility, fiber affinity) and comprise electron-donating groups such as amino and hydroxyl groups and electron-withdrawing groups such as nitro, carboxylic acid, and sulfonic acid groups.

References

1. Shore, J.: Historical development and classification of colorants. In: Shore, J. (ed.) *Colorants and Auxiliaries*, Vol. 1: Colorants, Society of Dyers and Colourists, pp. 5–6. Elsevier, London (1990)
2. Gürses, A., Açıkyıldız, M., Güneş, K., Gürses, M.S.: *Dyes and Pigments Springer Briefs in Molecular Science*, pp. 13–29. Springer, Cham (2016)
3. Abrahart, E.N.: *Dyes and Their Intermediates*, pp. 1–12. Chemical Publishing, New York (1977)
4. Zollinger, H.: *Color Chemistry*, p. 19. Wiley-VCH, Cambridge, UK (2003)

Chromostereopsis

Akiyoshi Kitaoka
Department of Psychology, Ritsumeikan
University, Kyoto, Japan

Synonyms

Color stereoscopic effect; Color stereoscopy;
Depth in color

Definition

Chromostereopsis refers to a phenomenon of binocular stereopsis that depends on binocular disparity due to the difference in color [1–4]. For example, red and blue on the same surface can appear to lie in different depth planes.

Overview

When the background is black (Fig. 1), the majority of observers see red objects closer to them than blue ones (red-in-front-of-blue observers), while a minority of observers see the reversal (blue-in-front-of-red observers) and the rest do not experience the phenomenon. On the other hand, when the background is white (Fig. 2), the perceived



Chromostereopsis, Fig. 1 Chromostereopsis. For the majority of observers, the *inset* made up of *red* random dots appears to be in front of the surround consisting of *blue* ones. For a minority, the *inset* appears to be behind the surround. The rest do not experience the phenomenon. Chromostereopsis is strong when observers see the image at a distance



Chromostereopsis, Fig. 2 When the background is *white*, the effect is reversed and it is weaker than when the background is *black*. Those who see *red* in front of *blue* in Fig. 1 see *blue* in front in this image, while those who see *blue* in front of *red* in Fig. 1 see *red* in front in this image

depth order is reversed, and the effect is weaker than when the background is black.

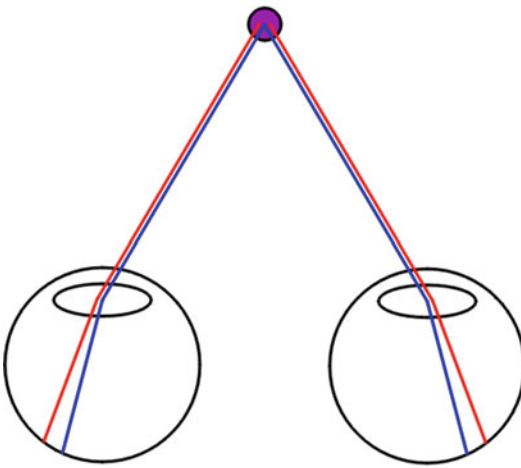
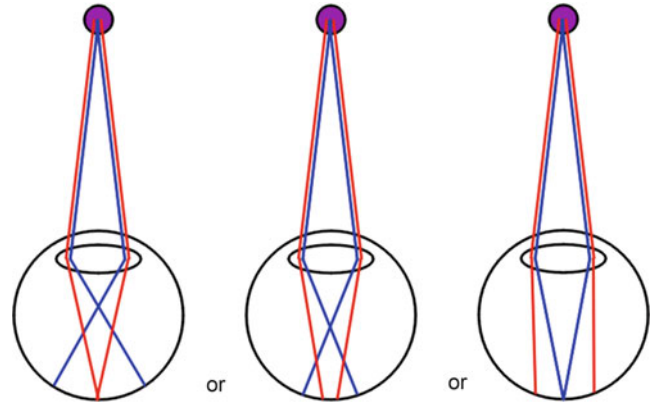
Goethe [5] may deserve to be the first person who proposed the notion of advancing color (red) versus receding color (blue), but he did not notice the binocular aspect of chromostereopsis. According to Thompson et al.'s review [6], the history of chromostereopsis goes back at least to the work of Donders in 1864 [7], while Vos' review [4] suggested that Donders first reported this effect in 1868, followed by Bruecke [8] in the same year. According to Dengler and Nitschke's review [9], however, Brewster [10] reported in 1851 that owing to the chromatic aberration of the lens, short-wavelength colors are seen stereoscopically as more distant.

Both Donders and Bruecke attributed chromostereopsis to accommodative feeling, which was translated to the perception of distance [4]. Blue or short-wavelength light has a larger refractive index than the red or long-wavelength one. This makes the color rays run in a different way as shown in Fig. 3. This is called "longitudinal chromatic aberration" [3]. This idea suggests that red should be closer than blue even if both colors are placed in the same depth plane because a closer object comes into focus at the posterior part. Moreover, this idea suggests that chromostereopsis should be seen monocularly. Yet, these two suggestions are not the cases, and this idea is not regarded as a plausible explanation of chromostereopsis [3, 11].

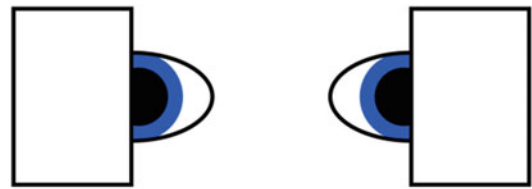
It was Bruecke [8] in 1868 who found the binocular nature of the phenomenon [4]. The optical axis of an eyeball is slightly (about 5° = angle gamma or angle alpha) shifted in the outward direction from the visual axis. Red light is thus projected to a more temporal part of the retina than does blue light (Fig. 4) because of the difference in refractive index of both colors. This physiological optics is called "transverse chromatic aberration" [3]. This idea suggests that red should be perceived in front of blue through binocular stereopsis based upon the binocular disparities of both colors. More specifically, a closer object projects to a more temporal part in the retina and so does red light. This suggestion, however, made

Chromostereopsis,

Fig. 3 The longitudinal chromatic aberration. *Blue* has the focus nearer to the lens than *red* because of the difference in the refractive index depending on color



Chromostereopsis, Fig. 4 The transverse chromatic aberration. *Red* light is projected to a more temporal part of the retina than does *blue* light because of the difference in the refractive index depending on color. Note that the optical axes are slightly (about 5°) diverged from the visual axes



Appearance: *blue* in front



Appearance: *red* in front

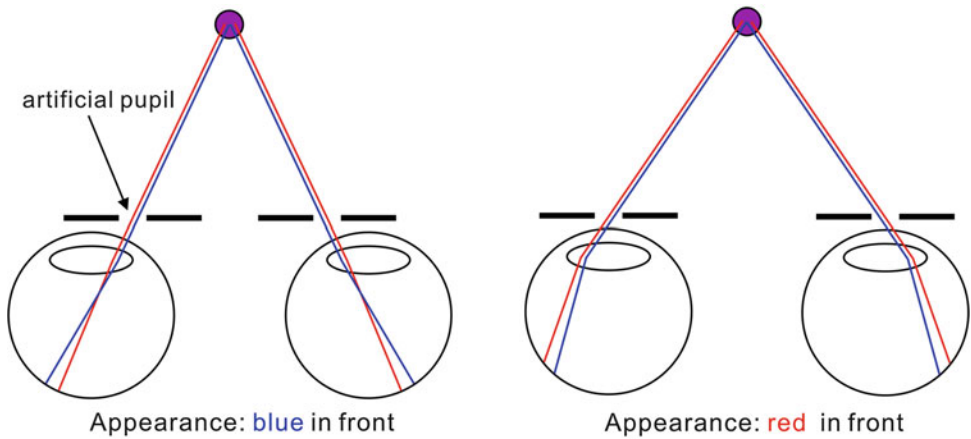
Chromostereopsis, Fig. 5 Einthoven's covering method. *Blue* in front of *red* is generated or enhanced by covering the temporal parts of both pupils (*upper image*), while *red* in front of *blue* is produced or enhanced by covering the nasal parts of both pupils (*lower image*)

Bruecke immediately reject his hypothesis because some of his observers reported red receding with respect to blue.

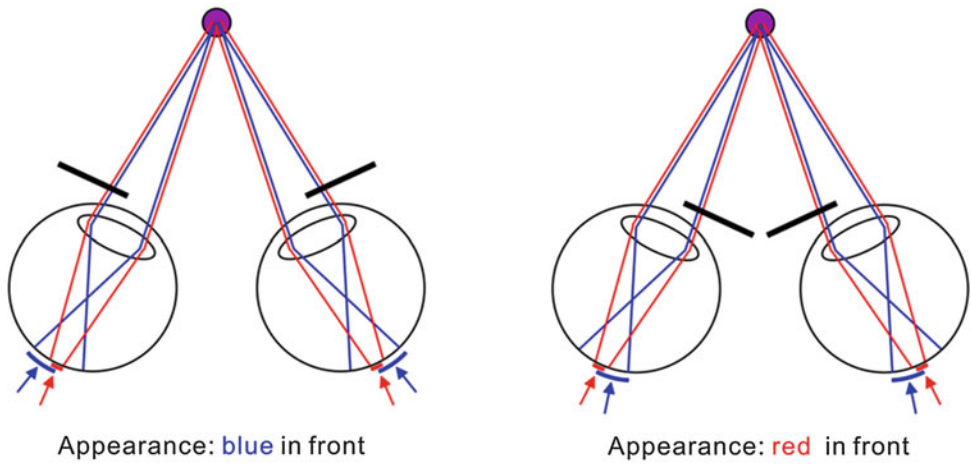
According to Vos [4], Einthoven discovered in 1885 that chromostereopsis is enhanced or reversed by using a simple method as shown in Fig. 5 [11, 12]. Covering the temporal parts of both pupils forces observers to see blue in front of red (upper image of Fig. 5). On the other hand, covering the nasal parts of both pupils makes observers see red in front of blue (lower image of Fig. 5). This method was also demonstrated by Kishto [13]. He cited Kohler's article [14] in

1962, which did not mention any other literature, though.

Subsequently, Einthoven's covering method was simplified to the method using artificial pupils. Nasally placed artificial pupils gave blue in front of red (upper image of Fig. 6), while temporally placed ones rendered red in front of blue (lower image of Fig. 6) [15–23]. Vos [4, 16, 18] attributed chromostereopsis to interactions between each individual pupil decentralization (angle gamma) and the Stiles-Crawford effect. The Stiles-Crawford effect is a phenomenon that the rays entering the eye through the peripheral



Chromostereopsis, Fig. 6 When artificial pupils are placed nasally, observers see *blue* in front of *red*. On the other hand, when artificial pupils are placed temporally, observers see *red* in front of *blue*



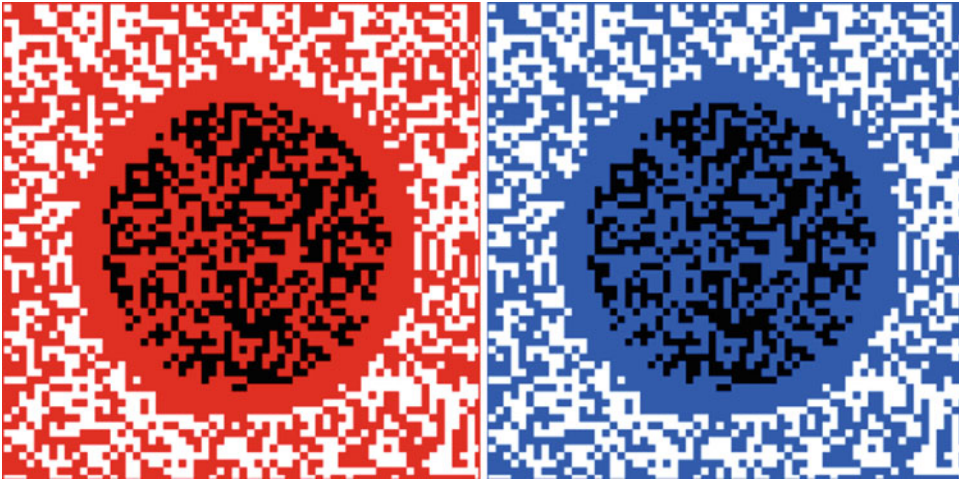
Chromostereopsis, Fig. 7 The center-of-gravity model. This model takes longitudinal chromatic aberration into account, in which the perceived position of a color is judged to be at the center of gravity of its diffused light

regions of the pupil are less efficient than those through the central region [24]. This two-factor model, which Vos [4] called the “generalized Bruecke-Einhoven explanation,” has been widely accepted, while a few authors did not approve it [25].

Many studies suggested that pupil size affects chromostereopsis [19, 21–23], which supports the generalized Bruecke-Einhoven explanation. Simonet and Campbell [26], however, did not find any consistent relationship between pupil size and chromostereopsis.

Einhoven’s original finding was explained by the center-of-gravity model (Fig. 7) [27]. It is

hypothesized that the position of color is determined at the center of gravity in the range of each projected light onto the retina. For example, when the temporal half is occluded, the center of gravity of red light shifts in the relatively nasal direction, while that of blue light deviates in the temporal direction (upper image of Fig. 7). These shifts give binocular disparities to produce the blue-in-front-of-red stereopsis. When the nasal half is occluded, the center of gravity of red light shifts in the relatively temporal direction, while that of blue light deviates in the nasal direction (lower image of Fig. 7). These shifts give binocular



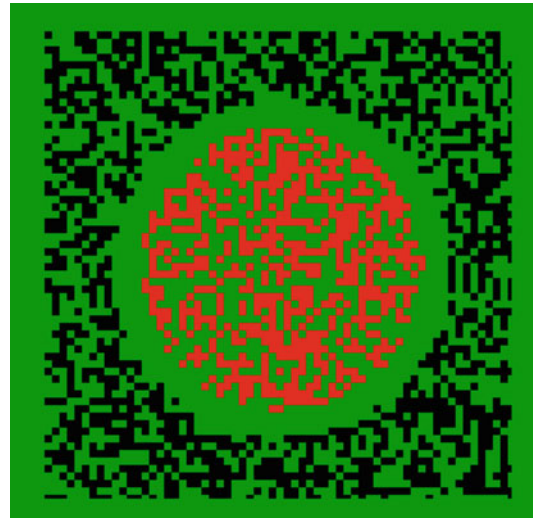
Chromostereopsis, Fig. 8 Images suggested by Hartridge [32]. *Red-in-front-of-blue* observers see the *black inset* in front of the *white* surround in the *left*

image, and they see the *inset* behind the surround in the *right image*. *Blue-in-front-of-red* observers see the reversal. Observe these images at a distance

disparities to produce the red-in-front-of-blue stereopsis.

In 1965, Kishto reported a tendency that red appears to be in front of blue at high levels of illumination, while blue appears to be in front of red at low levels of illumination, i.e., 17 of 25 observers (68 %) reported so [13]. This finding was influential and drew much attention [3], though it was questioned by some studies [9, 26]. For example, Simonet and Campbell [26] reported a reversal in the direction of the chromostereopsis for 16 of 30 observers (53%) when the ambient illumination was increased, but six of them (38%) reported a change toward the blue-in-front-of-red chromostereopsis. Moreover, at low illuminance, there was lack of correlation between the direction of chromostereopsis and transverse chromatic aberration, which may indicate that there may be a supplementary binocular factor in chromostereopsis [26].

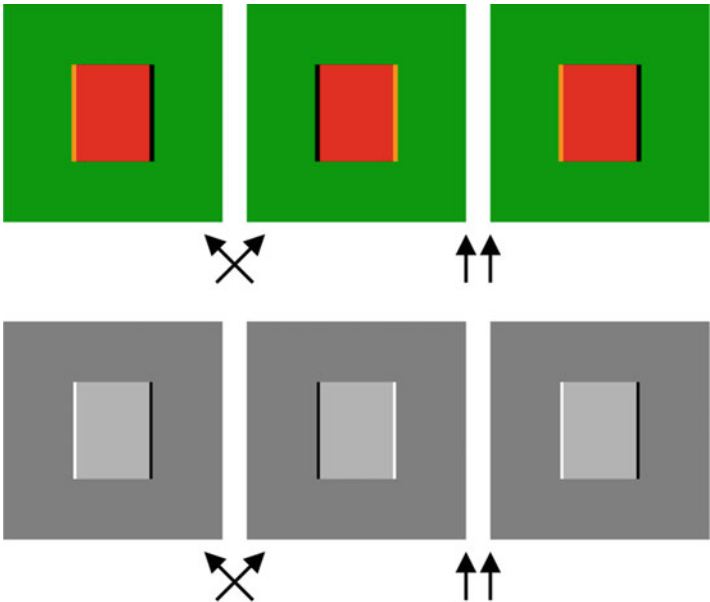
In 1928, Verhoeff reported that the perceived depth order between red and blue is reversed by changing the background from black to white (Fig. 2) [9, 28–30]. One account is that red surrounded by white lacks blue, while blue surrounded by white lacks red, suggesting that there are virtually blue and red edges, respectively [28]. According to Faubert [31], Hartridge described in his 1950s textbook [32] that “when both black and white pattern lie on a uniformly



Chromostereopsis, Fig. 9 Chromostereopsis when the background is not achromatic. Luminance of *red* is highest, followed by *green* in this image. Those who see *red* in front of *blue* in Fig. 1 see the *inset* in front of the surround; those who see *blue* in front of *red* in Fig. 1 see the *inset* behind the surround. Observe this image at a distance. Einthoven’s covering method (Fig. 5) also works

coloured background a stereoscopic effect is frequently noticed” (Fig. 8).

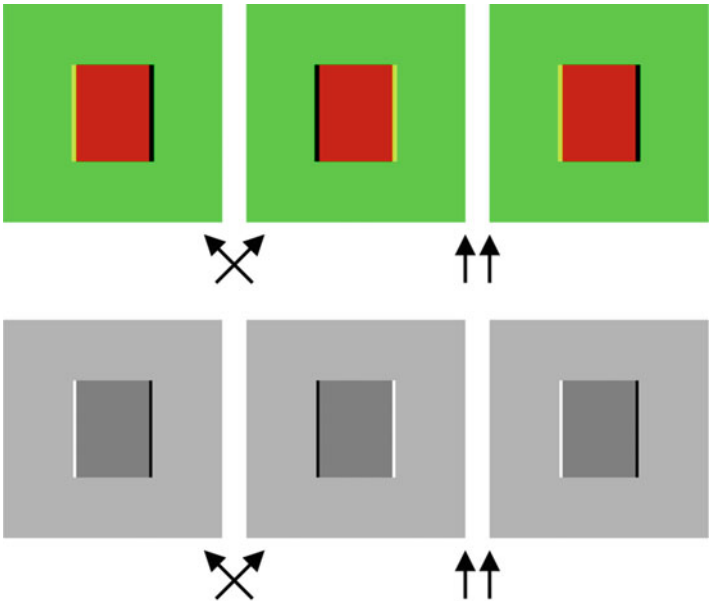
With respect to this luminance-dependent reversal, Faubert [31, 33] proposed a new demonstration of chromostereopsis in which colors are bordered with each other (Fig. 9) and pointed out



Chromostereopsis, Fig. 10 Faubert’s luminance-profile-based idea [31]. Color fringes or luminance profiles of a *red* object in front of *green* such as that in Fig. 9 produced by transverse chromatic aberration are depicted. It is supposed that mirror images are rendered to each eye. This binocular disparity generates a special type of binocular stereopsis which depends on luminance contrast polarity of the object’s edges [34, 35]. This stereogram

demonstrates *red-in-front-of-green* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones. The perceived depth is determined by the luminance profiles shown in the *lower row*. Note that *red* or *light-gray* rectangles do not give binocular disparity with respect to the frames; both fringes of each rectangle promote to make binocular stereopsis

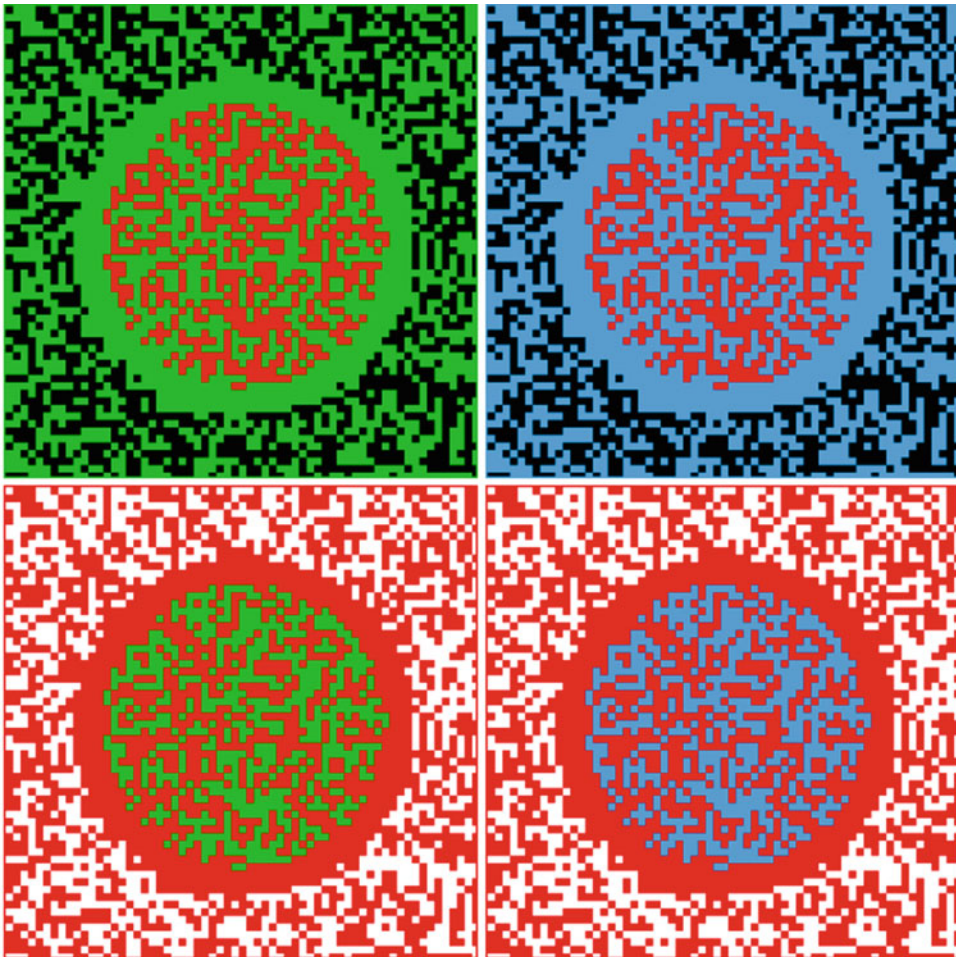
Chromostereopsis, Fig. 11 If the luminance order of the two colors in Fig. 10 is exchanged, the apparent depth is reversed. This stereogram demonstrates *red-behind-green* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones



the critical role of luminance profiles caused by transverse chromatic aberration in subsequent binocular stereopsis. The luminance-profile-dependent binocular stereopsis is thought to correspond to the one which Gregory and Heard showed in 1983 as shown in Fig. 10 [34, 35], though Faubert did not mention it. If the luminance order is reversed between the two colors, the apparent depth is reversed as shown in Fig. 11. When two colors are isoluminant, two depth planes are simultaneously observed (rivaldepth) with luster where two colors meet (Fig. 12) [31]. In addition, actually a century ago, Einthoven [11] had pointed out the role of luminance profiles caused by transverse chromatic aberration, but

he had assumed the perception of convex or concave objects depending on where illuminated light comes from and made a monocular explanation like the crater illusion (Fig. 13).

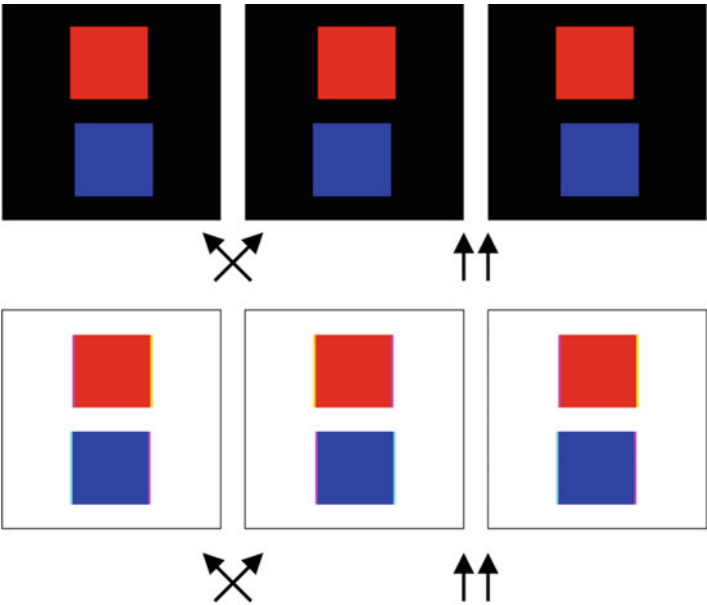
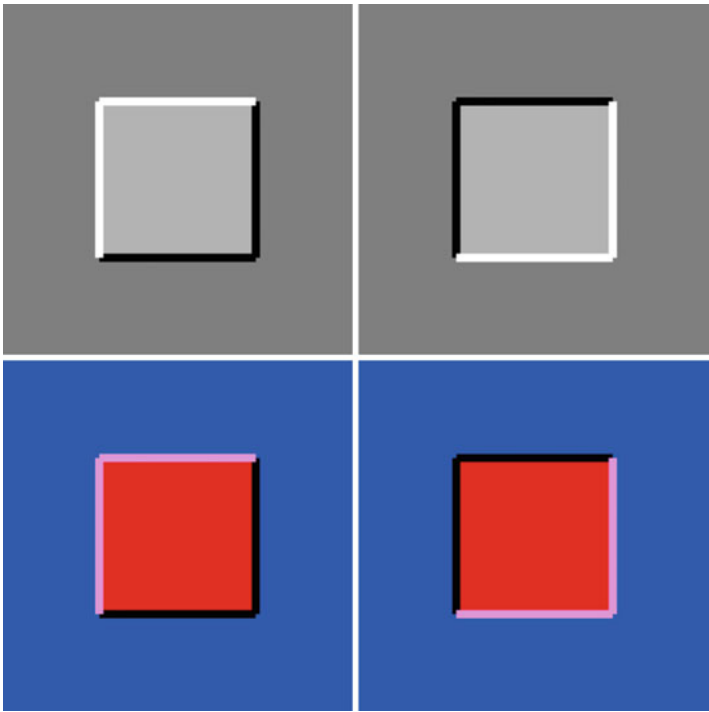
Faubert's luminance-profile-based model [31, 33] can be extended to explain the luminance-dependent reversal (Fig. 14). Suppose that white consists of red, green, and blue. Given transverse chromatic aberration when the background is black, and suppose that red appears to shift in one direction, blue appears to shift in the opposite direction (upper panel of Fig. 14), and green does not change the apparent position. Then, when the background is white, red appears to slightly shift in the opposite direction, with



Chromostereopsis, Fig. 12 Isoluminant colors make observers perceive rivaldepth with luster. Note that precise isoluminance is realized depending on displays and individuals

Chromostereopsis,

Fig. 13 Idea of the crater illusion. The central *square* appears to be in front of the background in the *upper left panel*, while that appears to be behind the background in the *upper right panel*. This depth effect depends on the positions of highlighted or shadowed edges [36]. The basic idea of Einthoven [11] is demonstrated in the *lower panels*

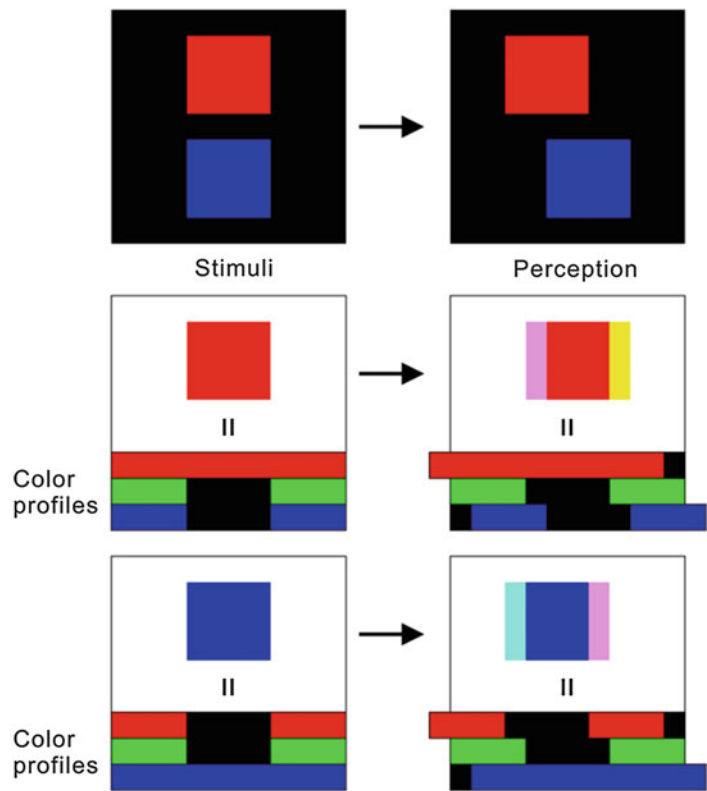


Chromostereopsis, Fig. 14 Faubert’s model can be extended to explain the color reversal. If *red* and *blue* objects are vertically aligned with the *black* background but appear to be shifted in position by transverse chromatic aberration as shown in the *upper panel*, the expected shifts with the *white* background are reversed as shown in the

lower panel. This idea accounts for the luminance-dependent reversal (Fig. 2). Note that this stereogram demonstrates *red-in-front-of-blue* appearance when cross-fusers use the *left* and *middle* panels or uncross-fusers see the *middle* and *right* ones

Chromostereopsis,

Fig. 15 If transverse chromatic aberration moves *red* to the *left* and *blue* to the *right* (uppermost row) and does not change the position of *green*, *red* surrounded by *white* appears to shift rightward (middle row), while *blue* surrounded by *white* appears to shift leftward (lowermost row). Thus, apparent position shifts of the two colors are reversed depending on whether the background is *black* or *white*. Note that the *left* column shows the physical position of the two colors, and the *right* column demonstrates apparent positions of the two colors with color fringes produced by color mixture of shifted component colors



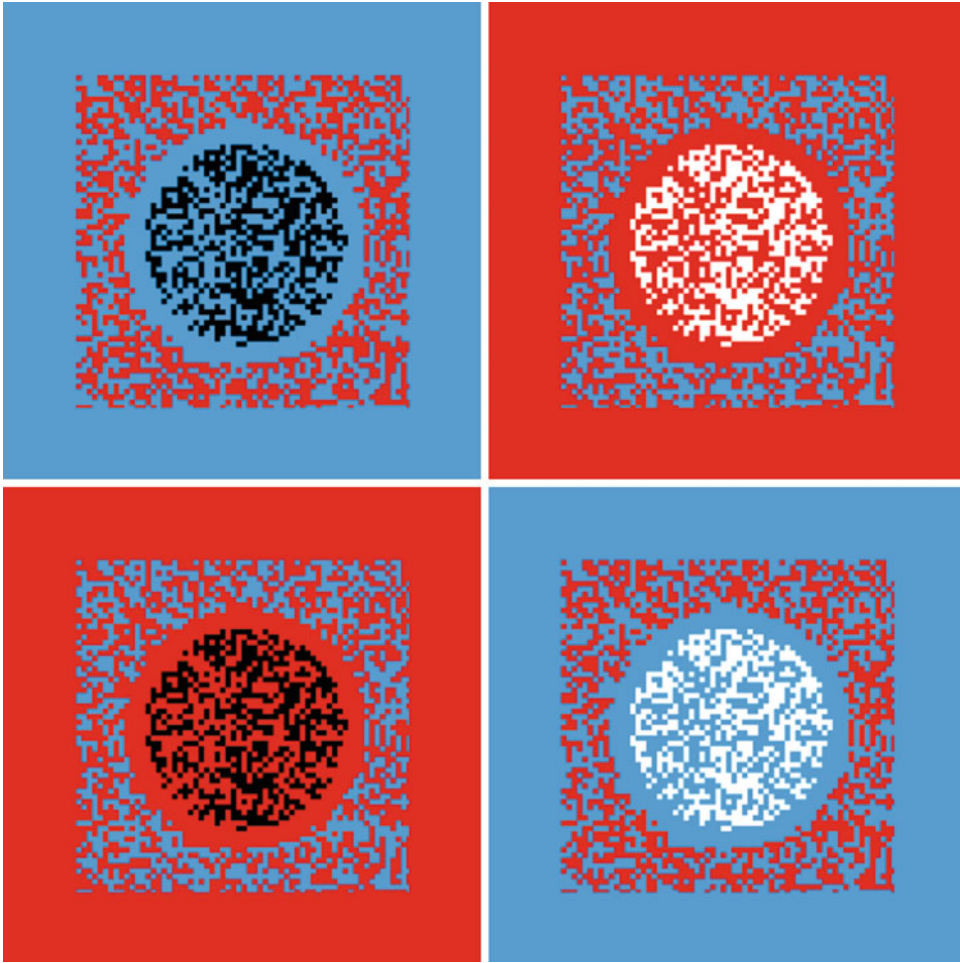
yellow leading to and magenta (color mixture of red and blue) following red. On the other hand, blue appears to slightly shift in the same direction as that of red when the background is black, with cyan (color mixture of blue and green) leading to and magenta following blue (lower panel of Fig. 14). Figure 15 shows the pictorial explanation of how transverse chromatic aberration induces the positional shifts of colors.

It was reported that the effect of chromostereopsis is large when observers see the image at a distance [13, 27, 31], whether observers are of the red-in-front-of-blue type or the blue-in-front-of-red type [27]. This issue remains open. In addition, there is no literature to suggest any involvement of myopia or hyperopia in chromostereopsis.

The majority of observers see red in front of blue with the black background in a light environment. How many people see blue in front of red? In Luckiesh [15], 11% (1 of 9 observers) did



Chromostereopsis, Fig. 16 The *inset* appears to be in front of the surround. High-luminance and/or high-contrast objects appear to be closer than low-luminance and/or low-contrast ones



Chromostereopsis, Fig. 17 Images showing a tendency that the *inset* appears to be behind the surround

so. The proportion was 4% (1 of 25 observers) in Kishto [13], 30% (9 of 30 observers) in Simonet and Campbell [26], 7% or 14% (1 or 2 of 14 observers) in Dengler and Nitschke [9], 20% (4 of 20 observers) in Kitaoka et al. [27], and 21% (16 of 75 observers) in Kitaoka [37].

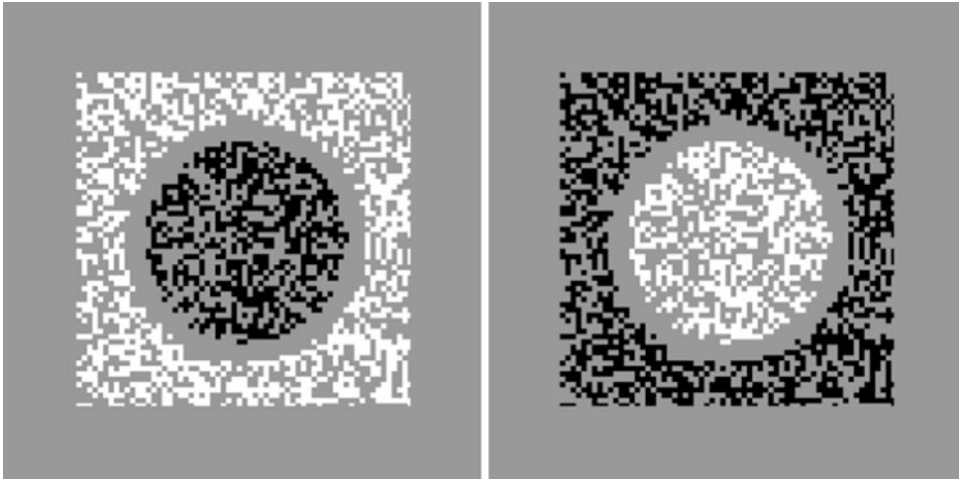
Color anomaly people also see chromostereopsis. Kishto [13] examined three color anomaly observers, one being a strong protanope, one being a mild deutanope, and the third having too poor color discrimination to read any of Ishihara plates. They all saw red in front of blue with the black background in a light environment.

It is thought that some part of the effect is due to luminance differences or contrast differences

(Fig. 16), with bright objects appearing closer than dim ones [6] or high-contrast objects appearing closer than low-contrast ones [38]. Saturation also affects chromostereopsis [11, 13]. Desaturation decreased the depth effect, though desaturation is inevitably accompanied by changes in luminance, contrast, or spectrum.

Moreover, if an image consists of the inset and surround, there seems to be a tendency that the inset appears to be behind the surround (Fig. 17). This phenomenon seems to be observed with achromatic images, too (Fig. 18).

Neural correlates of chromostereopsis were investigated using visually evoked potentials [39]. Results demonstrate that the depth illusion



Chromostereopsis, Fig. 18 Achromatic images also show the tendency that the *inset* appears to be behind the surround

obtained from contrast of color implicates similar cortical areas as classic binocular depth perception.

Summary

It is summarized that chromostereopsis is a phenomenon of binocular stereopsis based upon binocular disparity produced by some interaction between optic properties of color rays and further neural processing. Chromostereopsis is a ubiquitous phenomenon and has a considerable pile of observations gained in its long research history.

Cross-References

► [Stereo and Anaglyph Images](#)

References

1. Hartridge, H.: Chromatic aberration and resolving power of the eye. *J. Physiol.* **52**, 175–246 (1918)
2. Sundet, J.M.: Effects of colour on perceived depth: review of experiments and evaluation of theories. *Scand. J. Psychol.* **19**, 133–143 (1978)
3. Howard, I.P., Rogers, B.J.: *Binocular Vision and Stereopsis*. Oxford University Press, New York/Oxford (1995)
4. Vos, J.J.: Depth in colour, a history of a chapter in physiologie optique amusante. *Clin. Exp. Optom.* **91**, 139–147 (2008)
5. von Goethe, J.W.: *Zur Farbenlehre*. J. G. Cotta'sche Buchhandlung, Tübingen (1810)
6. Thompson, P., May, K., Stone, R.: Chromostereopsis: a multicomponent depth effect? *Displays.* **14**, 227–234 (1993)
7. Donders, F.C.: *On the Anomalies of Accommodation and Refraction of the Eye*, pp. 179–188. The New Sydenham Society, London (1864)
8. Brücke, E.: Über asymmetrische Strahlenbrechung im menschlichen Auge. *Sitz. Ber. Kaiserl. Akad. Wiss. Math. Natw. Klasse.* **II(58)**, 321–328 (1868)
9. Dengler, M., Nitschke, W.: Color stereopsis: a model for depth reversals based on border contrast. *Percept. Psychophys.* **53**, 150–156 (1993)
10. Brewster, D.: Notice of a chromatic stereoscope. *Philos. Mag., 4th Series.*, **3**, 31 (1851)
11. Einthoven, W.: On the prediction of shadow and perspective effects by difference of colour. *Brain.* **16**, 191–202 (1893)
12. Einthoven, W.: *Stereoscopie door kleurverschil*. PhD Thesis, Utrecht University (1885)
13. Kishto, B.N.: The colour stereoscopic effect. *Vision Res.* **5**, 313–329 (1965)
14. Kohler, I.: Experiments with goggles. *Sci. Am.* **206**, 69–72 (1962)
15. Luckish, M.: On “retiring” and “advancing” colors. *Am. J. Psychol.* **29**, 182–186 (1918)
16. Vos, J.J.: Some new aspects of color stereoscopy. *J. Opt. Soc. Am.* **50**, 785–790 (1960)
17. Vos, J.J.: An antagonistic effect in colour stereoscopy. *Ophthalmologica.* **145**, 442–445 (1963)
18. Vos, J.J.: The color stereoscopic effect. *Vision Res.* **6**, 105–107 (1966)

19. Sundet, J.M.: The effect of pupil size variations on the colour stereoscopic phenomenon. *Vision Res.* **12**, 1027–1032 (1972)
20. Sundet, J.M.: Two theories of colour stereoscopy. An experimental investigation. *Vision Res.* **16**, 469–472 (1976)
21. Owens, D.A., Leibowitz, H.W.: Chromostereopsis with small pupils. *J. Opt. Soc. Am.* **65**, 358–359 (1975)
22. Ye, M., Bradley, A., Thibos, L.N., Zhang, X.: Interocular differences in transverse chromatic aberration determine chromostereopsis for small pupils. *Vision Res.* **31**, 1787–1796 (1991)
23. Ye, M., Bradley, A., Thibos, L.N., Zhang, X.: The effect of pupil size on chromostereopsis and chromatic diplopia: interaction between the Stiles-Crawford effect and chromatic aberrations. *Vision Res.* **32**, 2121–2128 (1992)
24. Stiles, W.S., Crawford, B.H.: The luminous efficiency of rays entering the eye at different points. *Proc. R. Soc. Lond.* **112B**, 428–450 (1933)
25. Bodé, D.D.: Chromostereopsis and chromatic dispersion. *Am. J. Optom. Physiol. Opt.* **63**, 859–866 (1986)
26. Simonet, P., Campbell, M.C.: Effect of illuminance on the directions of chromostereopsis and transverse chromatic aberration observed with natural pupils. *Ophthalmic Physiol. Opt.* **10**, 271–279 (1990)
27. Kitaoka, A., Kuriki, I., Ashida, H.: The center-of-gravity model of chromostereopsis. *Ritsumeikan J. Human Sci.* **11**, 59–64 (2006)
28. Verhoeff, F.H.: An optical illusion due to chromatic aberration. *Am. J. Ophthalmol.* **11**, 898–900 (1928)
29. Hartridge, H.: The visual perception of fine detail. *Phil. Trans. R. Soc.* **232**, 519–671 (1947)
30. Winn, B., Bradley, A., Strang, N.C., McGraw, P.V., Thibos, L.N.: Reversals of the optical depth illusion explained by ocular chromatic aberration. *Vision Res.* **35**, 2675–2684 (1995)
31. Faubert, J.: Seeing depth in colour: more than just what meets the eyes. *Vision Res.* **34**, 1165–1186 (1994)
32. Hartridge, H.: *Recent Advances in the Physiology of Vision.* J & A Churchill, London (1950)
33. Faubert, J.: Colour induced stereopsis in images with achromatic information and only one other colour. *Vision Res.* **35**, 3161–3167 (1995)
34. Gregory, R.L., Heard, P.F.: Visual dissociations of movement, position, and stereo depth: some phenomenal phenomena. *Q. J. Exp. Psychol.* **35A**, 217–237 (1983)
35. Kitaoka, A.: Configurational coincidence among six phenomena: a comment on van Lier and Csathó (2006). *Perception.* **35**, 799–806 (2006)
36. Adams, W.J., Graf, E.W., Ernst, M.O.: Experience can change the “light-from-above” prior. *Nat. Neurosci.* **7**, 1057–1058 (2004)
37. Kitaoka, A.: A new development of studies on advancing and receding colors: modified longitudinal chromatic aberration model. KAKEN report. **15330159**, 1–45 (2007)
38. Mount, G.E., Case, H.W., Sanderson, J.W., Brenner, R.: Distance judgments of colored objects. *J. Gen. Psychol.* **55**, 207–214 (1956)
39. Cauquil, A.S., Delaux, S., Lestringant, R., Taylor, M.J., Trotter, Y.: Neural correlates of chromostereopsis: an evoked potential study. *Neuropsychologia.* **47**, 2677–2681 (2009)

Chromotherapy

► Color Psychology

CI Mordant Red 11

► Alizarin

CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments

János Schanda
Veszprém, Hungary

Synonyms

CIE Color-Matching Functions

Definition

CIE color-matching functions	Functions
	$\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ in the CIE 1931 standard colorimetric system or $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$ in the CIE 1964 standard colorimetric system

János Schanda: Deceased.

CIE 1931 standard colorimetric observer	Ideal observer whose color-matching properties correspond to the CIE color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ adopted by the CIE in 1931
CIE 1964 standard colorimetric observer	Ideal observer whose color-matching properties correspond to the CIE color-matching functions $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ adopted by the CIE in 1964

the additive mixture of the three matching primary stimuli (it is practical to use monochromatic red, green, and blue primary lights, see later) is projected onto the other side of the field. By using adjustable light attenuators, the light flux of the three matching stimuli is adjusted to obtain a color appearance match between the two fields. When this situation is reached, the test stimulus can be characterized by the three luminance values of the matching stimuli reaching the eye of the observer. This is described by the following equation:

$$[C] \equiv R[R] + G[G] + B[B] \quad (1)$$

where $[C]$ is the test stimulus; $[R]$, $[G]$, and $[B]$ are the unit values of the three matching stimuli; and R , G , and B are the amount taken from $[R]$, $[G]$, and $[B]$ to match the stimulus $[C]$. The $\equiv$ sign means “matches” (for more details, see [2]). Invariably, one of the primary stimuli has to be added to the test stimulus to complete the match.

If one uses monochromatic test stimuli of equal power, wavelength to wavelength along the visible spectrum (theoretically between 360 nm and 830 nm, practically between 380 nm and 780 nm), and matches these monochromatic radiations with the three selected matching stimuli (in different wavelength regions, one of the matching stimuli has to be added to the test stimulus to get a color match, and in that case the matching stimulus is shown in Eq. 1 as it would be subtracted from the other two matching stimuli), one can build up the color-matching functions.

Using $1 \text{ cd} \cdot \text{m}^{-2}$ monochromatic red light of 700 nm as $[R]$, $4.5907 \text{ cd} \cdot \text{m}^{-2}$ monochromatic green light of 546.1 nm as $[G]$ and $0.0601 \text{ cd} \cdot \text{m}^{-2}$ monochromatic blue light of 435.8 nm as $[B]$ for the units of the matching stimuli, one gets for unit power of monochromatic lights of the spectrum curves as shown in Fig. 1 called color-matching functions (CMFs) and usually written in the following form: $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$.

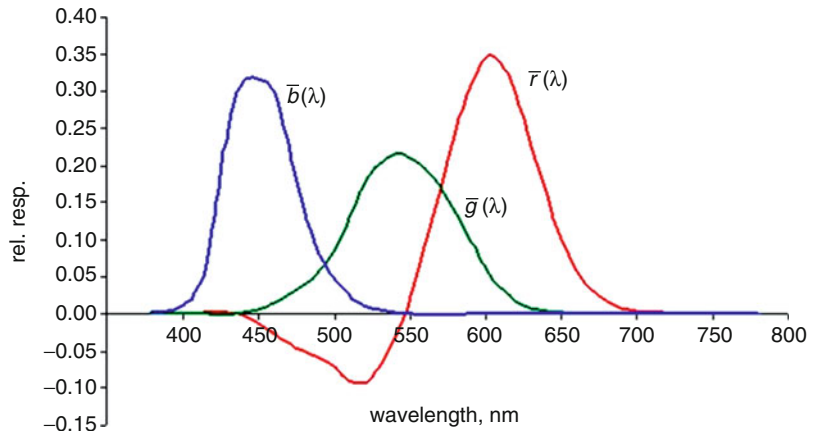
Visual experiments have shown that color stimuli are additive, i.e., if the test stimulus is composed of two sub-stimuli $[C(\lambda_1)]$ and $[C(\lambda_2)]$ of different wavelengths, the amounts of the $[R]$, $[G]$, and $[B]$ matching stimuli, also called *primaries*, that are used to match $[C(\lambda_1)]$ and

Overview

Colors with different spectral compositions can look alike (i.e., be metameric). An important function of colorimetry is to determine which spectral compositions appear metameric. The use of visual colorimeters for this purpose is handicapped by variations in the color matches made among observers classified as having normal color vision, so that different observers make different matches. Visual colorimetry also tends to be time-consuming. For these reasons, it has long been the practice in colorimetry to make use of sets of mean or standard color-matching functions to calculate tristimulus values for colors: Equality of tristimulus values for a pair of colors indicates that the color appearances of the two colors should match, when viewed under the same conditions by an observer for whom the color-matching functions apply. The use of standard sets of color-matching functions makes the comparison of tristimulus values obtained at different times and locations possible [1]. *The standard colorimetric observers are defined by their color-matching functions.*

According to Grassman’s laws, a *test color stimulus* can be matched by the additive mixture of three properly selected *matching primary stimuli* (properly selected includes independence; i.e., none of the primary stimuli can be matched by the additive mixture of the other two). The test stimulus is projected on one side of a bipartite field;

**CIE 1931 and 1964
Standard Colorimetric
Observers: History, Data,
and Recent Assessments,
Fig. 1** Color-matching
functions of the CIE 1931
RGB system



$[C(\lambda_2)]$ have to be added to match the additive mixture of the two test stimuli $[C_1 + {}_2]$. This can be expanded to a spectrum that has different spectral radiance values at different wavelength: the color of the compound spectrum $S(\lambda)$ can be described by three tristimulus values:

$$\begin{aligned} R &= \int_{380\text{nm}}^{780\text{nm}} \bar{r}(\lambda) S(\lambda) d\lambda, \\ G &= \int_{380\text{nm}}^{780\text{nm}} \bar{g}(\lambda) S(\lambda) d\lambda, \\ B &= \int_{380\text{nm}}^{780\text{nm}} \bar{b}(\lambda) S(\lambda) d\lambda \end{aligned} \quad (2)$$

In many applications, it is inconvenient to use negative lobes of the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ functions; therefore, the CIE decided in 1931 to transform the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ functions using a matrix transformation to imaginary CMFs (non-real in the sense that they cannot be physically realized). The transformed CMFs are the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ functions, and the tristimulus values determined using these functions are the X , Y , and Z tristimulus values. Their calculation is similar at those of the R , G , and B values shown in Eq. 2.

The transformation matrix has the following form:

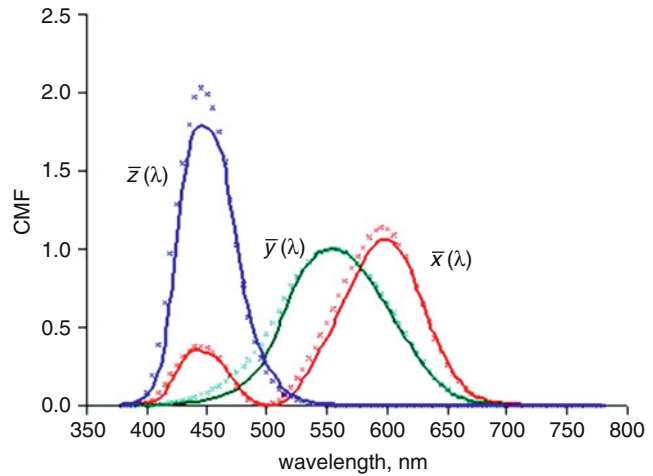
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.768 & 892 & 1.751 & 748 & 1.130 & 160 \\ 1.000 & 000 & 4.590 & 700 & 0.060 & 100 \\ 0 & 0.056 & 508 & 5.594 & 292 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

The original color-matching experiments were conducted with small, approximately 2° diameter homogeneous color patches, seen foveally. The central part of the fovea is covered by a yellow pigmentation, the yellow spot or macula lutea. If larger-colored fields are viewed or slightly off-axis objects are viewed, the above CMFs do not hold anymore, as the yellow pigmentation absorbs light in a part of the visible spectrum.

In 1964 CIE standardized CMFs for a 10° visual field, the so-called large field CMFs. Based on similar direct visual investigations performed mainly by Stiles and Burch [3] with contributions by Speranskaya [4], these functions are now used in many applications. To distinguish between values determined using the 2° or the 10° functions, the latter are distinguished by the subscript “10,” and thus the 10° observer’s CMFs are termed $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ (Fig. 2).

Regarding the use of the tristimulus values and further colorimetric calculations, see chapters on “CIE chromaticity co-ordinates,” “CIE chromaticity diagram,” “CIE illuminants,” and “CIELAB” and other chapters on advanced colorimetry.

CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments, Fig. 2 The $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ color-matching functions of the CIE 1931 standard (2°) colorimetric observer (full lines) and the $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ CMFs of the CIE 1964 standard (10°) colorimetric observer (shown by dashed line)



Short History of the CIE Colorimetric Observer

CIE colorimetry is based on the tristimulus theory developed by the greatest scientists of the nineteenth century, including Thomas Young, Helmholtz, and Maxwell (see [5]). Maxwell's demonstrations and ideas, in particular, lead to the specification of the trichromatic theory; he showed, e.g., the three color mixture curves of the spectrum and plotted the spectrum locus in the color triangle.

Photometry and colorimetry were further developed in the USA that lead to the so-called OSA excitation curves.

During the second decade of the twentieth century, two groups in the UK performed detailed investigations of color matching: John Guild at the NPL and David Wright at the Imperial College, London. The two researchers used different primaries, and it was a great surprise that after their transformation into a common system, they matched reasonably well.

During those years the CIE formulated the wish to develop a colorimetric description of colored glasses used for traffic control. At the 1931 meetings of the CIE, the USA and UK groups discussed in detail the pros and cons of different systems and finally agreed that a mean of the Guild and Wright data should be adopted, but transformed to a system with nonnegative

CMFs. These were the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ CMFs that we still use today.

One major problem with the CIE 1931 XYZ system is that the values of the primaries were determined photometrically, so that the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ functions had to be reconstructed using $V(\lambda)$ function, the visibility function (now called spectral luminous efficiency for photopic vision). As it turned out later, the $V(\lambda)$ dataset is in error in the blue part of the spectrum, and this error has been transferred to the color-matching functions.

Experiments carried out in the 1950s at NPL by Stiles and Burch led to a new set of 10-deg CMFs; this time the primaries and test lights were radiometrically calibrated, so they were not contaminated by photometric errors. As mentioned in the Overview, the NPL data were harmonized with data measured by Speranskaya, and these were standardized as the CIE 10° observer in 1964.

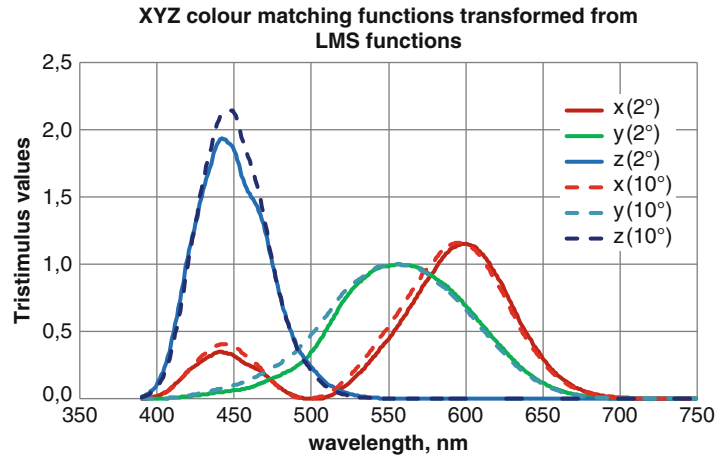
Recently, the fundamental experimental data have been reanalyzed, together with new cone spectral sensitivity measurements made in red-green dichromats and blue-cone monochromats of known genotype to produce new cone spectral sensitivity curves of the three cone receptors (the cones are mainly responsible for daylight vision and color perception) as well as transformations to $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ -like curves [6, 7]; see Chapter on CIE Physiologically-Based Colour-Matching Functions and Chromaticity Diagrams. The cone fundamentals are known as the

CIE 1931 and 1964

Standard Colorimetric

Observers: History, Data, and Recent Assessments,

Fig. 3 The $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$, and $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ color-matching functions transformed from the LMS cone fundamentals



fundamental CMFs or $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$. Both the 2° and the 10° CMFs have been published by the Colour & Vision Research Laboratory at <http://www.cvgl.org> (see Fig. 3) – at the time of submitting this manuscript, CIE TC 1-36 has not endorsed the transformation from the LMS cone fundamentals to the CMFs shown here.

Schanda, J. (ed.) CIE colorimetry – understanding the CIE system. Wiley Interscience, Hoboken (2007)

6. CIE: Fundamental Chromaticity Diagram with Physiological Axes – Part 1. CIE Publ. 170-1:2006
7. CIE: Fundamental Chromaticity Diagram with Physiological Axes – Part 2. under publication

Cross-References

- CIE Chromaticity Coordinates (xyY)
- CIE Chromaticity Diagrams, CIE purity, CIE dominant wavelength
- CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams
- CIELAB
- Daylight Illuminants

References

1. CIE Colorimetry – Part 1: CIE standard colorimetric observers. CIES014-1/E:2006 – ISO 11664-1:2007(E)
2. Schanda, J.: Chapter 3: CIE colorimetry. In: Schanda, J. (ed.) CIE Colorimetry – Understanding the CIE System. Wiley Interscience, Hoboken (2007)
3. Stiles, W.S., Burch, J.M.: N.P.L. colour-matching investigation: Final report (1958). Opt. Acta. **6**, 1–26 (1959)
4. Speranskaya, N.I.: Determination of spectral color co-ordinates for twenty-seven normal observers. Opt. Spectr. **7**, 424–428 (1959)
5. Wright, W.D.: Chapter 2: The history and experimental background to the 1931 CIE system of colorimetry. In:

CIE 1976 L*a*b*

- CIELAB

CIE 1976 L*u*v* Color Space

- CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space

CIE 1994 (ΔL^* ΔC^*_{ab} ΔH^*_{ab})

- CIE94, History, Use, and Performance

CIE 2000 Color-Difference Equation

- CIEDE2000, History, Use, and Performance

CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02

Ming Ronnier Luo

School of Design, University of Leeds, Leeds, UK
University of Zhejiang, Zhejiang, China

Synonyms

CAT

Definition

According to the CIE International Lighting vocabulary [1], chromatic adaptation is a visual process whereby approximate compensation is made for changes in the colors of stimuli, especially in the case of changes in illuminants. The effect can be predicted by a chromatic adaptation transform (CAT) which is used to predict the corresponding colors, a pair of color stimuli that have same color appearance when one is seen under one illuminant and the other is seen under the other illuminant.

Overview

CAT is used for many industrial applications. For example, it is highly desired to produce color constant merchandise, i.e., products do not change color appearance across different illuminants. A color inconstancy index named CMCCON02 was proposed by the Colour Measurement Committee (CMC) of the Society of Dyers and colorists (SDC) [2]. CAT is a key element in the color inconstancy index. It later became the ISO standard for textile applications [3]. Furthermore, chromatic adaptation is the most important function included in a color appearance model, which is capable of predicting color appearance under different viewing conditions such as illuminants, levels of luminance, background colors, media (e.g., reflection,

transmissive, and self-luminous display). The CIE [4] recommended CIECAM02 for the application of the color management systems. For illumination engineering, a CAT is also required for predicting the color rendering properties between a test and a reference illuminant [5].

Various chromatic adaptation transforms have been derived to fit a particular experimental dataset (see later). The workflow for a typical CAT is given in Fig. 1 including three steps.

Step 1 Cone response transform

To model the physiological mechanisms of chromatic adaptation, one must express stimuli in terms of cone responses, denoted R , G , and B or L , M , and S , suggestive of long-wave (red), middle-wave (green), and short-wave (blue) sensitivities, respectively. This is achieved by using a linear transform via a 3 by 3 matrix. Various transform functions have been proposed having different fundamental primaries.

Step 2 Chromatic adaptation mechanism

This step transforms cone responses of the test sample, (R, G, B) , under the test illuminant, defined by (R_w, G_w, B_w) , into the adapted cone responses (R_c, G_c, B_c) under the reference illuminant, defined by (R_{rw}, G_{rw}, B_{rw}) . The transforms are different between different CATs.

Step 3 Reverse cone transformation

Using the reverse cone transform (an inverse matrix of Step 1) to calculate the corresponding cone responses, (R_c, G_c, B_c) in Step 2), back to tristimulus values under the reference illuminant.

CIE TC1-52 technical report entitled *A Review of Chromatic Adaptation Transformations* [6] gave a comprehensive survey of the transforms, and reported the testing results of the state of the art CATs using large accumulation of experimental datasets.

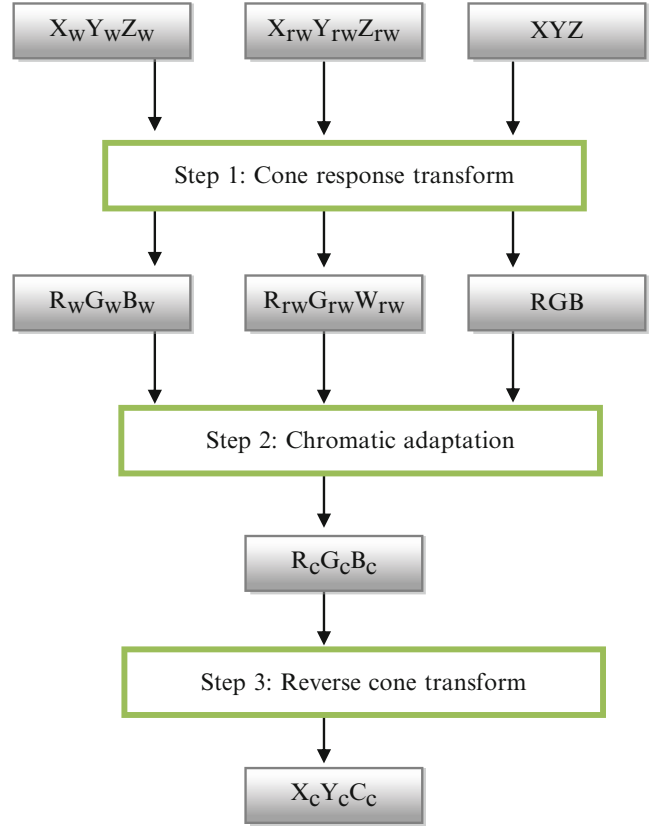
Four of them, the most well-known, are introduced below. The notation used in each CAT is different from those used in its original version, but agree with those given in Fig. 1.

von Kries Chromatic Adaptation Transform

The earlier CAT was that developed by von Kries in 1902. He studied chromatic adaptation

CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02,

Fig. 1 The three steps included in a chromatic adaptation transform



following the Young-Helmholtz theory, which assumes that, although the responses of the three cone types (R , G , B) are affected differently by chromatic adaptation, the relative spectral sensitivities of each of the three cone mechanisms remain unchanged. Hence, chromatic adaptation can be considered as a change of sensitivity by a constant factor for each of the three cone mechanisms. The magnitude of each factor depends upon the color of the stimulus to which the observer is adapted. The relationship, given in Eq. 1, is known as the *von Kries coefficient law*.

$$\begin{aligned} R_c &= \alpha R \\ G_c &= \beta G \\ B_c &= \gamma B \end{aligned} \quad (1)$$

where R_c , G_c , B_c and R , G , B are the cone responses to give the same appearance under the

reference and test illuminants respectively. The α , β , and γ are the von Kries coefficients corresponding to the change in sensitivity of the three cone mechanisms due to chromatic adaptation. These are calculated using Eq. 2.

$$\alpha = \left(\frac{R_{rw}}{R_w} \right); \quad \beta = \left(\frac{G_{rw}}{G_w} \right); \quad \gamma = \left(\frac{B_{rw}}{B_w} \right) \quad (2)$$

where

$$\frac{R}{R_w} = \frac{R_c}{R_{rw}}; \quad \frac{G}{G_w} = \frac{G_c}{G_{rw}}; \quad \frac{B}{B_w} = \frac{B_c}{B_{rw}}$$

and R_{rw} , G_{rw} , B_{rw} and R_w , G_w , B_w are the cone responses for the reference white under the reference and test illuminants, respectively.

In 1974, the CIE technical committee on color rendering [5] adopted a version of the von Kries model derived by Helson et al. [7]. It is still in use for making small adjustments to account for

differences in illuminants to be compared for color rendering properties. This procedure is given below.

Step 1 Calculate the R , G , B , R_{rw} , G_{rw} , B_{rw} , and R_w , G_w , B_w using Judd's cone transformation in Eq. 3.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0,000 & 1,000 & 0,000 \\ -0,460 & 1,360 & 0,100 \\ 0,000 & 0,000 & 1,000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

Step 2 Calculate the α , β , and γ von Kries coefficients and the R_c , G_c , B_c values using Eqs. 1 and 2.

Step 3 Calculate the X_c , Y_c , Z_c using Eq. 4.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} 2,954 & -2,174 & 0,220 \\ 1,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 1,000 \end{bmatrix} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \quad (4)$$

CIELAB

Although the CIELAB color space [8] was recommended by CIE in 1976 solely for quantifying color differences under daylight illuminants, it can also be used with other illuminants because it includes a von Kries type of transformation, i.e., by dividing the tristimulus values (X , Y , Z) of the sample by those (X_w , Y_w , Z_w) of illuminant, respectively. The assumption made is that L^* , a^* , and b^* values will be the same for a pair of color constant colors under a test and a reference illuminant.

CMCCAT97 Chromatic Adaptation Transformation

Lam and Rigg [9] investigated the degree of color appearance change for object colors between the test and reference illuminants. They conducted a memory matching experiment using 58 textile samples under illuminants D65 and A. A transformation was derived to fit the experimental data. The transform was named BFD transform, which

is similar to the structure of Bartleson's. At a later stage, this transform was enhanced by Luo and Hunt [10] to become CMCCAT97. It was also included in the first version of CIE color appearance model, CIECAM97s [11]. CMCCAT97 transform is given below.

A parameter, L_a (adapting luminance), is required for calculating a parameter, D , which allows for the degree of chromatic adaptation taking place. L_a is calculated as $L_w Y_b / 100$ where L_w is the luminance in cd/m^2 of the reference white under test illuminant and Y_b is the luminance factor of the background. The whites are normally the perfect reflecting diffuser, in which case $Y_w = Y_{rw} = 100$. Other whites may be used but, to avoid ambiguity, their details should be recorded.

Step 1 Calculation of R , G , B , R_{rw} , G_{rw} , B_{rw} and R_w , G_w , B_w using Eq. 5.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M \begin{bmatrix} X/Y \\ Y/Y \\ Z/Y \end{bmatrix}$$

where

$$M = \begin{bmatrix} 0.8951 & 0.2664 & -0.1614 \\ -0.7502 & 1.7135 & 0.0367 \\ 0.0389 & -0.0685 & 1.0296 \end{bmatrix} \quad (5)$$

Step 2 Calculation of degree of adaptation (D) using Eq. 6.

$$D = F - \frac{F}{1 + 2L_a^{1/4} + L_a^2/300} \quad (6)$$

where $F = 1$ for surface samples seen under typical viewing conditions.

D is set to one by assuming that the color of the illuminant is usually discounted during visual color-inconstancy assessments for object colors. This is proposed by CMCCON97 [12].

Step 3 Calculation of the corresponding RGB cone responses using Eq. 7.

$$\begin{aligned} R_C &= [D(R_{wr}/R_w) + 1 - D]R \\ G_C &= [D(G_{wr}/G_w) + 1 - D]G \\ B_C &= [D(B_{wr}/B_w^p) + 1 - D]|B|^p \end{aligned} \quad (7)$$

(when B is negative, B_c must be made negative)

$$\text{where } p = (B_w/B_{wr})^{0.0834}$$

Step 4 Calculation of the corresponding tristimulus values using Eq. 8.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = M^{-1} \begin{bmatrix} R_c Y \\ G_c Y \\ B_c Y \end{bmatrix} \quad (8)$$

CAT02 Transform

Various tries were carried out to test the performance of different transformations by CIE TC1-52 *Chromatic Adaptation Transformation*. At the same period, CIE TC 8-01 *Color Appearance Models for Color Management Applications* was aimed to improve the CIECAM97s model and in some degree to simplify the model. This resulted in CIECAM02 [4] which includes CAT02. At a later stage, CMC also modified its original CMCCON97 color constancy index [12] to CMCCON02 [2] by replacing CMCCAT97 by CAT02. The CAT02 is given below:

Step 1 Calculation of $R, G, B, R_{rw}, G_{rw}, B_{rw}$ and R_w, G_w, B_w using Eq. 9.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M_{CAT02} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (9)$$

where

$$M_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix}$$

Step 2 Calculation of degree of adaptation (D) using Eq. 10.

$$D = F \left[1 - \left(\frac{1}{3.6} \right) e^{\left(\frac{L_a - 42}{92} \right)} \right] \quad (10)$$

where F is set to 1.0, 0.9 or 0.8 for “average,” “dim,” or “dark” surround condition, respectively and L_a is the luminance of the adapting field. In theory D should range from 0 for no adaptation to the adopted white point to 1 for complete adaptation to the adopted white point. In practice the minimum D value will not be less than 0.65 for a “dark” surround and will exponentially converge to 1 for “average” surrounds. If D from Eq. 10 is larger than one, D should be set to one. For predicting color inconstancy of a sample using CMCCON02, D value should be set to one assuming a complete adaptation.

Step 3 Calculation of R_c, G_c, B_c from R, G, B (similarly R_{wc}, G_{wc}, B_{wc} from R_w, G_w, B_w)

$$\begin{aligned} R_c &= [D(R_{rw}/R_w) + 1 - D]R \\ G_c &= [D(G_{rw}/G_w) + 1 - D]G \\ B_c &= [D(B_{rw}/B_w) + 1 - D]B \end{aligned} \quad (11)$$

Step 4 Calculation of the corresponding tristimulus values using equation 12.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = M_{CAT02}^{-1} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \quad (12)$$

Experimental Datasets Investigated by CIE TC1-52

CIE TC1-52 members also collected large number of datasets. These were accumulated using mainly three different psychophysical methods [13]: Haploscopic matching, memory matching, and magnitude estimation. The former is to ask observers to perform color matching between two viewing fields say one eye views a stimulus under illuminant A and the other eye views another stimulus under illuminant D65. Memory matching is to ask observers to describe color stimuli using a particular color order system such as Munsell under one illuminant with full adaptation. Prior the experiment, the observers

CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02, Table 1 List of classical experiments for each technique

Viewing-field	Experiment	Year	References
Haploscopic matching			
Simple	CSAJ	1991	[14]
Complex	Breneman	1987	[15]
Memory matching			
Simple	Helson et al.	1952	[7]
Complex	Lam & Rigg	1985	[9]
	Braun & Fairchild	1996	[16]
Magnitude estimation			
Simple	Kuo & Luo	1995	[17]
Complex	Luo et al	1991	[[18], [19]]
	Luo et al	1993	[[20], [21]]

need to be trained to learn the perceptual attributes such as Munsell Value, Chroma, and Hue. Magnitude estimation is to ask observers to estimate color stimuli in terms of lightness, colorfulness, and hue. Table 1 shows the datasets chosen to be extensively studied by CIE TC1-52.

Evaluation of CATs

The four CATs introduced earlier were tested using the datasets given in Table 1 from eight sources. In total, 568 corresponding-color pairs were gathered. The results are summarized in Table 2 in terms of mean CMC(1:1) Δ*E* units [22].

It was found [17] that the typical observer variation for studying chromatic adaptation was about four CMC (1:1) units. Hence, if a CAT has an error of prediction equal to or less than four units, it may be considered to be satisfactory. As shown in Table 2, the most accurate transform for each dataset (the underlined and bold value) is usually less than four units except for Luo et al. (WF) data set. The ten data sets are divided into two groups: reflection and non-reflection samples. The Braun and Fairchild, and Breneman data sets are included in the latter group. (Braun and Fairchild data were obtained by comparing between CRT and reflection printed colors, and Breneman

data were based on projected transmissive colors.) The weighted mean for each CAT was calculated to represent the performance for each group or overall. The weighted mean was used to take into account the number of corresponding pairs in each data set.

The results showed that for both data groups, CMCCAT97 and CAT02 outperformed von Kries and CIELAB with a big margin, and the former and the latter two gave similar degree of error of prediction. This implies that the former two and the latter two CATs gave very similar results. This implies that von Kries law alone is insufficient to develop a reliable CAT. The matrix in Step 1 of Fig. 1 is essential for a reliable CAT.

Also, CMCCAT07 was derived by fitting only one dataset, the Lam and Rigg. It also predicts well to the other datasets. This implies that all corresponding datasets agree reasonably well with each other. CAT02 can be considered as an improvement of CMCCAT97 because it is simpler and was developed by fitting all the datasets in Table 2.

Another method to evaluate the performance of CATs is to visualize the color shifts in a color space. The predicted shifts for the four CATs in Table 2 and the corresponding experimental shifts from all data were plotted in CIELAB *a***b** diagram. The Helson et al. dataset [7] was used to illustrate the transformations to CIE C from A illuminant as shown in Fig. 2. This dataset was chosen because it has been the most widely used and the earlier produced dataset. In Fig. 2, the point where the two vectors cross, and the unmarked end represent experimental results viewed under illuminants A and C, respectively. The “+” symbol represents the predicted chromaticity from one particular transformation. The distance between each corresponding “+” and unmarked end indicates the error of prediction except for CIELAB. For a good agreement between experimental results and a particular transformation, the two vectors should overlap. For perfect agreement between CIELAB and the experimental results, each vector in Fig. 2a Reference source not found. should have a zero length (or a single point). As can be seen in Fig. 2, such perfect agreement was not found. However, there is a clear pattern of color shift in each figure. That



CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02, Table 2 The performance of chromatic-adaptation transforms

Data sets\Transform	Refer. Illum.	Test illum.	No. of Pairs	CIELAB	von Kries	CMC CAT97	CAT02
Group 1: Reflective							
CSAJ-C	D65	A	87	5.0	4.1	<i>3.4</i>	<i>3.4</i>
Helson et al.	C	A	59	6.2	5.1	<i>3.8</i>	<i>3.8</i>
Lam & Rigg	D65	A	58	5.0	5.0	<i>3.4</i>	3.6
Luo et al. (A)	D65	A	43	5.6	5.5	3.9	<i>3.8</i>
Luo et al. (D50)	D65	D50	44	4.8	<i>4.1</i>	4.2	4.2
Luo et al. (WF)	D65	WF	41	<i>4.5</i>	6.1	4.7	4.7
Kuo et al. (A)	D65	A	40	5.6	5.8	3.6	<i>3.5</i>
Kuo et al. (TL84)	D65	F11	41	3.3	3.9	2.8	<i>2.6</i>
Group 1 weighted mean				5.1	4.9	<i>3.7</i>	<i>3.7</i>
Group 2: Non-reflective							
Braun & Fairchild	D65	3000 K 9300 k	66	5.5	5.2	3.7	<i>3.6</i>
Breneman	D65	Various	107	8.2	8.0	5.6	<i>5.5</i>
Group 2 weighted mean				7.2	6.9	4.9	<i>4.8</i>
Overall weighted mean				<i>5.7</i>	<i>5.7</i>	<i>4.0</i>	3.9

Note: The bold and underlined italic values indicate the best transform for each data set

is, the color shift increases for more colorful colors as C_{ab}^* increases.

Note that experimental errors would be expected to be random. When a diagram shows a consistent pattern in the errors of prediction of a particular color region, this is most likely to be due to a fault in the transform. See the example the von Kries diagram at very colorful regions.

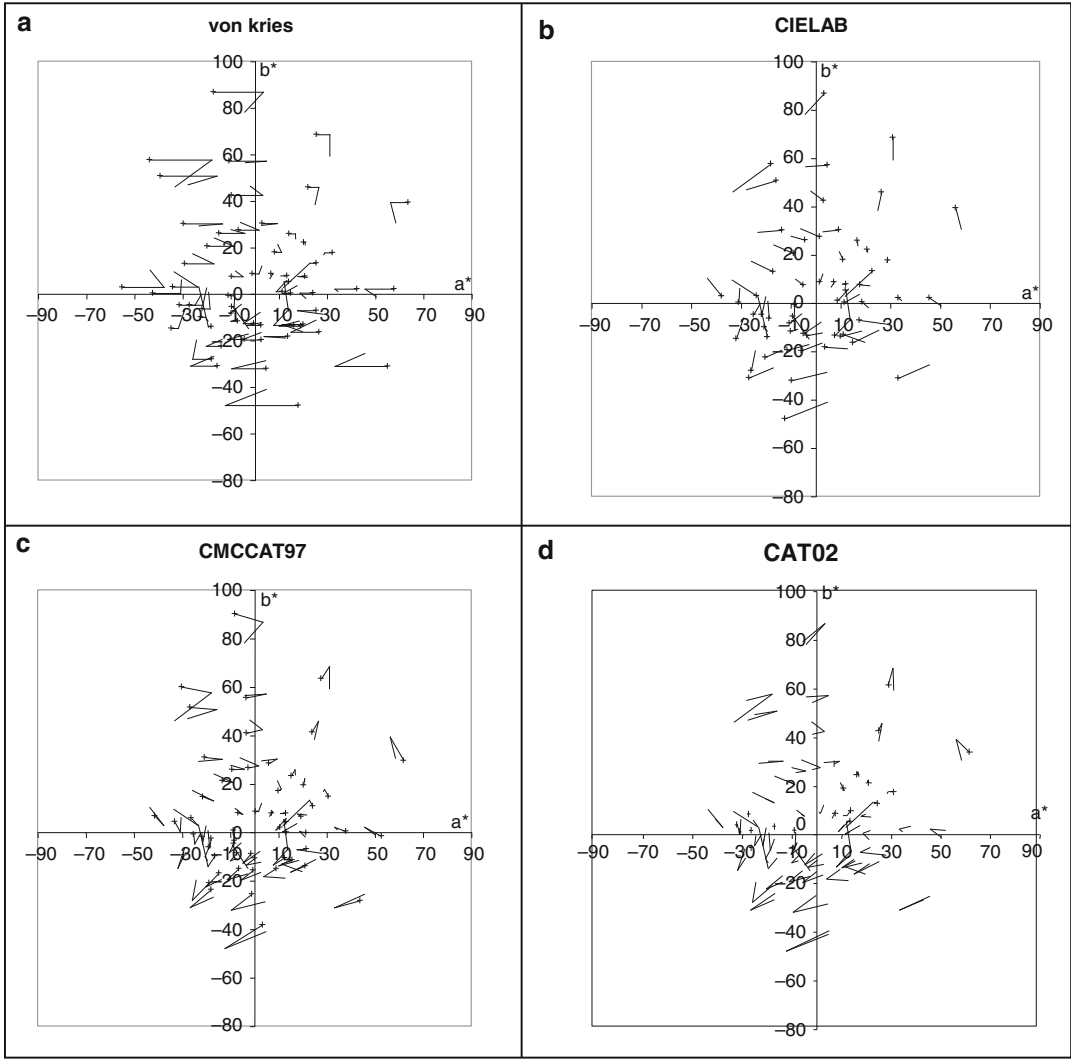
Figs. 2a–d show that there are large differences between the four different CATs in terms of predictive color shifts. For the von Kries transform (Fig. 2b), the predictive shifts only move along the a^* direction, i.e., red-green shift. Both von Kries and CIELAB gave reasonable predictions for the low chroma colors, but large predictive errors for high chroma colors. CMCCAT97 (Fig. 2c) gave a quite precise prediction for almost all colors with some exceptions in the colorful yellow and blue regions. The prediction of those regions were improved for CAT02 transform (Fig. 2d). It can also be seen that in general, the magnitudes and shifts for CAT02 are very similar to those of CMCCAT97 (see Fig. 2c). This indicates that although CMCCAT97 was derived to fit only the Lam and Rigg dataset, it gave almost the same performance as that CAT02 (see Table 2). This

implies that there is a great similarity between different datasets

Further Extension to CAT16

The CATs, especially CAT02, have been applied successfully in various applications. However, some shortcomings have been identified for some very saturated colors (close to the spectrum locus of the chromaticity diagram) [23]. These colors result in negative tristimulus values, meaning out of the limit of chromaticity diagram. Although they are rare in real applications, efforts have been made to correct them. In 2016, Li et al. [24] published CAM16, CAT16, and CAM16-UCS by modifying CIECAM02, CAT02, and CAM02-UCS, respectively. The main effort was to overcome the shortcoming by replacing M_{CAT02} in Eq. (9) by M_{16} matrix in Eq. (13) in CAT02.

$$M_{16} = \begin{pmatrix} 0.401288 & 0.650173 & -0.051461 \\ -0.250268 & 1.204414 & 0.045854 \\ -0.002079 & 0.048952 & 0.953127 \end{pmatrix} \quad (13)$$



CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02, Fig. 2 Graphical presentation of corresponding a^*b^* values showing direction and magnitude of the experimental visual results under CIE standard illuminant A and C, plotted between the point

where the two vectors cross and the unmarked end, and between the former point and the predicted shifts plotted using an “+” symbol, whereas (a) von Kries, (b) CIELAB, (c) CMCCAT97, and (d) CAT02

The difference between CAT16 and CAT02 is the latter to include negative lobes in the cone response functions, and the functions are narrower, and the latter to have all positive cone response functions with all positive energy across the visible spectrum. The two CATs gave exactly the same performance in CIELAB color difference to predict all corresponding datasets. It is envisaged that CAM16 and CAT16 will be

adopted by CIE to become CIECAM16 and CIECAT16 respectively in the near future.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [CIELAB](#)

References

1. CIE International Lighting vocabulary, Central Bureau of the CIE, Vienna (2012). It is available at website of <http://eilv.cie.co.at/>
2. Luo, M.R., Li, C.J., Hunt, R.W.G., Rigg, B., Smith, K.J.: The CMC 2002 colour inconstancy index: CMCCON02. *Color. Technol.* **119**, 280–285 (2003)
3. ISO 105-J05 Textiles: Test for colour fastness. Part 5 Method for the instrumental assessment of the colour inconstancy of a specimen with change in illuminant (CMCCON02)
4. CIE Pub. No. 159:2004 A Colour Appearance Model for Colour Management Systems: CIECAM02
5. CIE Pub. No. 13.2:1974 Method of measuring and specifying colour rendering properties of light sources
6. CIE Pub. No. 160:2004 A Review of Chromatic Adaptation Transforms
7. Helson, H., Judd, D.B., Warren, M.H.: Object-color changes from daylight to incandescent filament illumination. *Illum. Eng.* **47**, 221–233 (1952)
8. CIE Pub.15: Colorimetry. Central Bureau of the CIE, Vienna (2004)
9. Lam K.M.: Metamerism and Colour Constancy, Ph.D. Thesis, University of Bradford (1985)
10. Luo, M.R., Hunt, R.W.G.: A chromatic adaptation transform and a colour inconstancy index. *Color. Res. Appl.* **23**, 154–158 (1998)
11. Luo, M.R., Hunt, R.W.G.: The structure of the CIE 1997 colour appearance model (CIECAM97s). *Color. Res. Appl.* **23**, 138–146 (1998)
12. Luo, M.R., Hunt, R.W.G., Rigg, B., Smith, K.J.: Recommended colour inconstancy index. *J. Soc. Dyers. Col.* **115**, 183–188 (1999)
13. Luo, M.R., Rhodes, P.A.: Corresponding-colour data sets. *Color. Res. Appl.* **24**, 295–296 (1999)
14. Mori, L., Sobagaki, H., Komatsubara, H., Ikeda, K.: Field trials on CIE chromatic adaptation formula. In: *Proceedings of the CIE 22nd Session – Division 1*, pp. 55–58 (1991)
15. Breneman, E.J.: Corresponding chromaticities for different states of adaptation to complex visual fields. *J. Opt. Soc. Am. A* **A4**, 1115–1129 (1987)
16. Braun, K.M., Fairchild, M.D.: Psychophysical generation of matching images for cross-media colour reproduction. In: *Proceedings of the 4th Color Imaging Conference*, pp. 214–220. IS&T, Springfield, Va (1996)
17. Kuo, W., Luo, M.R., Bez, H.: Various chromatic adaptation transforms tested using new colour appearance data in textiles. *Color. Res. Appl.* **20**, 313–327 (1995)
18. Luo, M.R., Clarke, A.A., Rhodes, P.A., Scrivener, S.A.R., Schappo, A., Tait, C.J.T.: Quantifying colour appearance. Part I. LUTCHI colour appearance data. *Color. Res. Appl.* **16**, 166–180 (1991)
19. Luo, M.R., Clarke, A.A., Rhodes, P.A., Scrivener, S.A.R., Schappo, A., Tait, C.J.T.: Quantifying colour appearance. Part II. Testing colour models performance using LUTCHI colour appearance data. *Color. Res. Appl.* **16**, 181–197 (1991)
20. Luo, M.R., Gao, X.W., Rhodes, P.A., Xin, J.H., Clarke, A.A., Scrivener, S.A.R.: Quantifying colour appearance. Part III. Supplementary LUTCHI colour appearance data. *Color. Res. Appl.* **18**, 98–113 (1993)
21. Luo, M.R., Gao, X.W., Rhodes, P.A., Xin, J.H., Clarke, A.A., Scrivener, S.A.R.: Quantifying colour appearance. Part VI. Transmissive media. *Color. Res. Appl.* **18**, 191–209 (1993)
22. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dyers Col.* **100**, 128–132 and 281–282 (1984)
23. Li, C.J., Perales, E., Luo, M.R., Martinez-Verdu, F.: Mathematical approach for predicting non-negative tristimulus values using the CAT02 chromatic adaptation transform. *Color. Res. Appl.* **37**, 255–260 (2012)
24. Li, C.J., Li, Z., Wang, Z., Xu, Y., Luo, M.R., Cui, G., Melgosa, M., Brill, M., Pointer, M.R.: Comprehensive color solutions: CAM16, CAT16 and CAM16-UCS. *Color. Res. Appl.* **42**, 703–718 (2017)

CIE Chromaticity Coordinates (xyY)

Stephen Westland
Colour Science and Technology, University of
Leeds, Leeds, UK

Definition

Chromaticity coordinates x , y , and z are calculated from CIE tristimulus values X , Y , and Z thus:

$$\begin{aligned}x &= X/(X + Y + Z), \\y &= Y/(X + Y + Z), \\z &= Z/(X + Y + Z).\end{aligned}\tag{1}$$

If the chromaticity coordinates and one of the tristimulus values (e.g., Y) are known, then it is possible to compute tristimulus values thus:

$$\begin{aligned}X &= xY/y, \\Z &= zY/y = (1 - x - y)Y/y.\end{aligned}\tag{2}$$

Chromaticity Coordinates

The use of chromaticity coordinates is an alternative, and often more useful, representation to the use of tristimulus values. The tristimulus values X , Y , and Z of a stimulus are the amounts of three primary lights that on average an observer would use to match the stimulus and as such form a specification of the stimulus. If the Y tristimulus value is calculated on an absolute basis, then it represents the luminance of the color stimulus, in candelas per square meter, for example [1]. When relative tristimulus values are calculated so that $Y = 100$ for a similarly illuminated and viewed perfect Lambertian reflector, then the Y of a stimulus is equal to its reflectance factor or, in some cases, transmittance factor. It is customary to regard Y as the luminance factor of the stimulus, and this is an approximate correlate of the perceptual attribute of lightness [1].

The CIE system of colorimetry was designed so that the Y tristimulus value correlates, at least approximately, with lightness. To achieve this and other objectives, the primaries upon which the CIE system is based are usually referred to as being *imaginary* and certainly cannot be physically realized. This means that the X and Z tristimulus values do not correlate, even approximately, with any perceptual attributes, and this is a motivation for calculating other attributes that can provide such correlates. The chromaticity coordinates (see Eq. 1) are a type of relative tristimulus values. So, for example, if $X = 20$, $Y = 40$, and $Z = 20$ for a stimulus, then $x = 20/80 = 0.25$ and $y = 40/80 = 0.5$. This indicates that the stimulus is 25% of X , 50% of Y , and 25% of Z , of course.

Inherent in the way that the chromaticity coordinates are calculated (Eq. 1) is the constraint that $x + y + z = 1$, and this means that there are only really two degrees of freedom since $z = 1 - x - y$. Since there are only two *free* variables, it is possible to construct a 2D diagram referred to as a chromaticity diagram. By convention, in the chromaticity diagram (which forms a sort of map of colors), x and y are plotted on the abscissa and ordinate, respectively (Fig. 1).

Properties of the Chromaticity Diagram

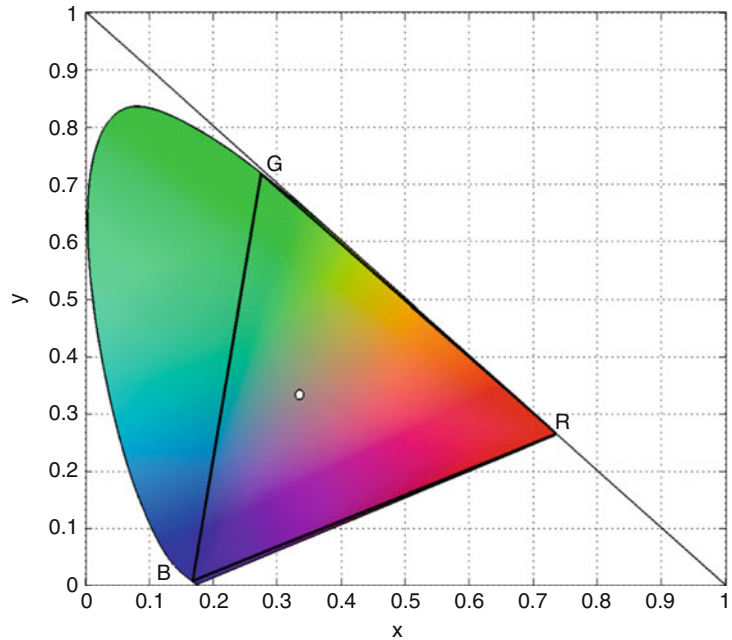
Figure 1 shows an illustration of a chromaticity diagram though note that the colors are purely representative and are not meant to accurately denote the color at any point in the diagram. The original color-matching experiments carried out by Wright and Guild that formed the basis of the CIE system of color specification in 1931 used real, but different, red (R), green (G), and blue (B) lights as the primaries but were transformed into a common set of monochromatic primaries; the R was at 700 nm, the G at 546.1 nm, and the B at 435.8 nm. The color-matching functions are the amounts of each of these primaries used on average by a group of observers to match each wavelength of light in the visible spectrum (380–780 nm). The vertices of the triangle in Fig. 1 are at the chromaticities of the CIE RGB primaries.

If two lights are represented in the chromaticity diagram by two points, then chromaticities of the additive mixtures of the two lights will be represented by the straight line that joins the two points. Thus, in Fig. 1, all mixtures of the R and G primaries would lie on the straight line joining the chromaticities of the R and G primaries. The range of colors that can be matched by a set of primaries is sometimes referred to as the gamut; the gamut of a dichromatic system (based on just two primaries) is very small and impractical for most purposes. When there are three primaries (a trichromatic system), then the gamut becomes a triangle in the CIE chromaticity diagram such as the RGB triangle illustrated in Fig. 1.

The curved horseshoe-shaped locus of the chromaticity diagram is defined by the chromaticities of the monochromatic wavelengths of light. Since all real color stimuli are combinations of the monochromatic wavelengths, and given the earlier observation about how color mixtures are defined in the chromaticity diagram, it is clear that the gamut of all physically realizable colors is the convex hull constrained by the curved spectral locus. Similarly, it is also clear that no matter how carefully three primaries are selected (and no matter whether they are monochromatic or not),

CIE Chromaticity Coordinates (xyY),

Fig. 1 CIE 1931 chromaticity diagram showing the equal-energy stimulus (white circle) and the CIE RGB gamut (triangle denoted by R, G, and B)



the gamut (represented by a triangle in the chromaticity diagram) will always be a subset of the gamut of all physically realizable colors. Thus, if the gamut of the CIE RGB primaries is considered (see Fig. 1), it is evident that much of the spectral locus cannot be matched by additive mixture of the primaries. The 1931 CIE system was defined by transforming the color-matching functions from the RGB primaries into a system of imaginary primaries XYZ where the whole spectrum (from 380 to 780 nm) could be matched by all-positive amounts of the three primaries. The XYZ primaries are referred to as imaginary because they cannot be realized physically; in Fig. 1, the outer triangle – defined by the x,y chromaticities (0,0), (1,0), and (0,1) – is the gamut of the XYZ primaries, and it is evident that the gamut of physically realizable colors lies within this.

The largest gamut that can be achieved by any RGB system would be obtained by using primaries whose chromaticities were close to, or on, the spectral locus at the approximate wavelengths 400, 520, and 700 nm, but this would still leave parts of the gamut of physically realizable colors outside of the RGB triangle (see Fig. 2). A further consideration is that a practical RGB system consists of chromaticity and luminance.

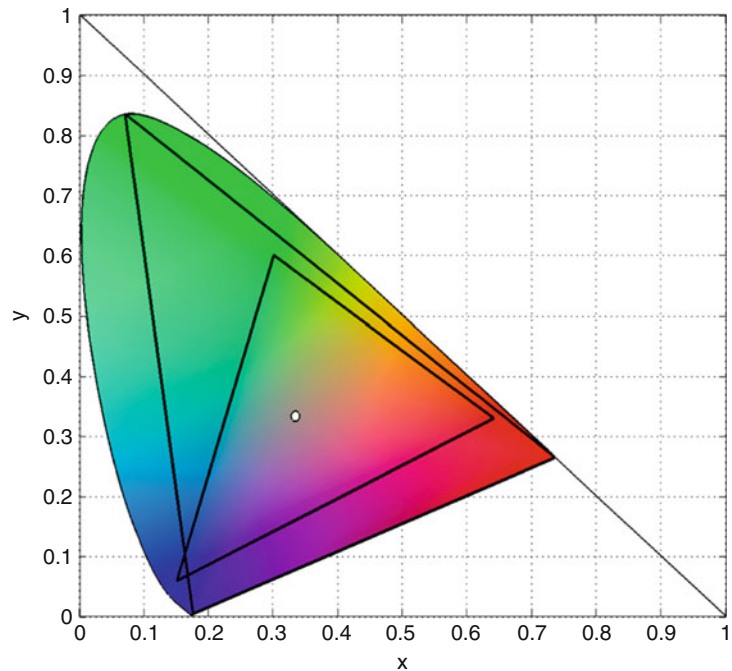
Consequently, an RGB display device based on monochromatic primaries would likely not be very bright. For many purposes, it is important to be aware that gamuts are three dimensional [2].

Primaries that consist of light at more than one wavelength are less saturated than monochromatic lights of a similar hue; however, they also tend to be much brighter. The design of modern display devices involves many such considerations, but many use primaries that correspond closely to the sRGB trichromatic standard whose gamut is represented in Fig. 2 [3]. It is evident that the sRGB gamut (in 2D chromaticity space at least) covers less than half of the gamut of physically realizable colors. However, the fact that reflectance spectra for objects in the world tend to vary smoothly with wavelength [4] has the consequence: the practical gamut of real-world colors is much smaller than the horseshoe-shaped locus would suggest. Monochromatic stimuli, for example, are incredibly rare in the natural or man-made world.

The chromaticity diagram is perceptually non-uniform [1]. This was visually demonstrated by the MacAdam ellipses which showed the chromaticities of stimuli that were just noticeably different from a standard color. Around each standard

CIE Chromaticity Coordinates (xyY),

Fig. 2 CIE chromaticity diagram showing the gamut of a hypothetical monochromatic RGB primary system (*larger triangle*) and the sRGB standard primary system (*smaller triangle*)



color, the locus of the just discriminable colors was the ellipses whose size and orientation varied greatly throughout the chromaticity diagram. In a perceptually uniform space, these loci would be circles of identical size. Even lines of constant hue are curved in chromaticity space rather than being straight. The Abney effect, first observed in 1909, is a phenomenon such that there is a hue shift when white light is added to a monochromatic light [5]. The locus of the mixture of a white light (see the equal-energy stimulus in Fig. 1) and a monochromatic light would be a straight line in the chromaticity diagram. Problems with the lack of perceptual uniformity of the chromaticity diagram were part of the reason why non-linear transforms of the XYZ system were explored ultimately resulting in the CIE (1976) $L^*a^*b^*$ color space or CIELAB.

Further Considerations and Future Directions

Currently, there are two CIE xy chromaticity spaces corresponding to the 1931 (2° of visual angle) and 1964 (10° of visual angle) standard

observers, respectively. Work is underway to explore the possibility of a CIE standard observer that would include a visual angle parameter to allow a family of related chromaticity diagrams.

Cross-References

- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Tristimulus Values](#)
- [CIELAB](#)

References

1. Hunt, R.W.G., Pointer, M.R.: *Measuring Colour*, 4th edn. Wiley, Hoboken (2011)
2. Morovič, J.: *Color Gamut Mapping*. Wiley, Hoboken (2008)
3. Westland, S., Cheung, V.: RGB systems. In: Chen, J., Cranton, W., Fihn, M. (eds.) *Handbook of Visual Display Technology*, pp. 147–154. Springer, Berlin (2012)
4. Maloney, L.: Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am.* **3**(10), 1673–1683 (1986)
5. Pridmore, R.: Effect of purity on hue (Abney effect) in various conditions. *Color. Res. Appl.* **32**(1), 25–39 (2007)

CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength

János Schanda
Veszprém, Hungary

Definitions

Chromaticity Diagram

Plane diagram in which points specified by chromaticity coordinates represent the chromaticities of color stimuli [1]

Note: In the CIE standard colorimetric systems, y is normally plotted as ordinate and x as abscissa, to obtain an x, y chromaticity diagram.

Purity (of a Color Stimulus)

Measure of the proportions of the amounts of the monochromatic stimulus and of the specified achromatic stimulus that, when additively mixed, match the color stimulus considered.

Note 1: In the case of purple stimuli, the monochromatic stimulus is replaced by a stimulus whose chromaticity is represented by a point on the purple boundary.

Note 2: The proportions can be measured in various ways.

Excitation Purity [p_e]

Quantity defined by the ratio NC/ND of two collinear distances on the chromaticity diagram of the CIE 1931 or 1964 standard colorimetric systems, the first distance being that between the point C representing the color stimulus considered and the point N representing the specified achromatic stimulus and the second distance being that between the point N and the point D on the spectrum locus at the dominant wavelength of the color stimulus considered, leading to the following expressions:

$$p_e = \frac{y - y_n}{y_d - y_n}$$

or

$$p_e = \frac{x - x_n}{x_d - x_n}$$

where (x, y) , (x_n, y_n) , (x_d, y_d) are the x, y chromaticity coordinates of the points C, N, and D, respectively.

Unit: 1

Note 1: In the case of purple stimuli, see Note 1 to “purity.”

Note 2: The formulae in x and y are equivalent, but greater precision is given by the formula which has the greater value in the numerator.

Note 3: Excitation purity, p_e , is related to colorimetric purity, p_c , by the equation:

$$p_e = \frac{p_c y}{y_d}$$

Colorimetric Purity [p_c]

Quantity defined by the relation

$$p_c = \frac{L_d}{L_n + L_d}$$

where L_d and L_n are the respective luminances of the monochromatic stimulus and of the specified achromatic stimulus that match the color stimulus considered in an additive mixture.

Note 1: In the case of purple stimuli, see Note 1 to “purity.”

Note 2: In the CIE 1931 standard colorimetric system, colorimetric purity, p_c , is related to excitation purity, p_e , by the equation $p_c = p_e y_d / y$ where y_d and y are the y chromaticity coordinates, respectively, of the monochromatic stimulus and the color stimulus considered.

Note 3: In the CIE 1964 standard colorimetric system, a measure, $p_{c,10}$, is defined by the relation given in Note 2, but using $p_{e,10}$, $y_{d,10}$, and y_{10} instead of p_e , y_d , and y , respectively.

Dominant Wavelength (of a Color Stimulus)
[λ_d]

Wavelength of the monochromatic stimulus that, when additively mixed in suitable proportions with the specified achromatic stimulus, matches the color stimulus considered in the CIE 1931 x, y chromaticity diagram.

Unit: nm

Note: In the case of purple stimuli, the dominant wavelength is replaced by the complementary wavelength.

Overview

x, y Chromaticity Diagram

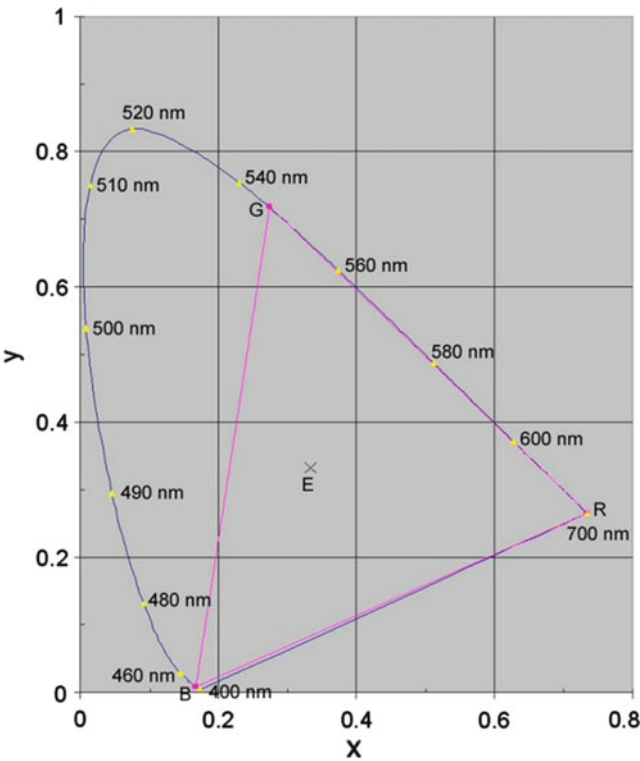
Both in the CIE 1931 standard colorimetric system and the CIE 1964 standard colorimetric system, chromaticity coordinates are expressed as the ratio of the given tristimulus value and the sum of all three tristimulus values [2]:

$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}. \quad (1)$$

As the color-matching functions are the tristimulus values of the monochromatic stimuli, the chromaticity coordinates of the monochromatic stimuli can be calculated according to Eq. 1. In the plane rectangular x - y diagram the line of the chromaticity of the monochromatic stimuli bounds, together with the straight line connecting the red and blue endpoints of the spectrum, the area of visible stimuli [3]; see Fig. 1. The diagram produced by plotting x as abscissa and y as ordinate is called the CIE 1931 chromaticity diagram or the CIE (x, y) diagram. A similar chromaticity diagram can be constructed using the x_{10}, y_{10} chromaticity coordinates of the CIE 1964 standard colorimetric system.

The chromaticity diagram is often depicted in color; see Fig. 3 (in the section for dominant wavelength and purity). One has to emphasize, however, that in this figure, the colors are only for orientation. As shown in Fig. 1, if, e.g., on the computer the colors are mixed from the R, G, B primaries (the gamut of real RGB primaries of monitors is even smaller), the mixed colors have

CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength, Fig. 1 CIE 1931 chromaticity diagram, location of the equi-energy spectrum (E), the R, G, B primaries of the CIE 1931 system, and some wavelength along the spectrum loci are shown



to be inside the RGB triangle. On the boundary of the monochromatic stimuli, the emission spectrum reaching our eyes should contain only one single wavelength [4].

As seen in Fig. 3, the mid-part of the chromaticity diagram looks whitish. This is even more pronounced if a light source of that chromaticity illuminates a scene; a white paper will – under these conditions – look white, and this is caused by chromatic adaptation.

Chromaticity diagrams can be built also for the CIE 1976 u' , v' coordinates [5]. The 1976 u' , v' uniform chromaticity scale diagram (UCS diagram) is a projective transformation of the CIE 1931 x , y chromaticity diagram yielding perceptually more uniform color spacing, i.e., the perceived chromaticity differences are represented by more uniform coordinate differences. The transformation between the two systems is

$$\begin{aligned} u' &= \frac{4x}{-2x + 12y + 3} \\ v' &= \frac{9y}{-2x + 12y + 3} \end{aligned} \quad (2)$$

With these coordinates, the chromaticity diagram has the form as shown in Fig. 2. Comparing this diagram with the x , y diagram, it becomes

obvious how nonuniform the x , y diagram is (see details in entry ► “CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space”).

An equivalent transformation starting from the tristimulus values is

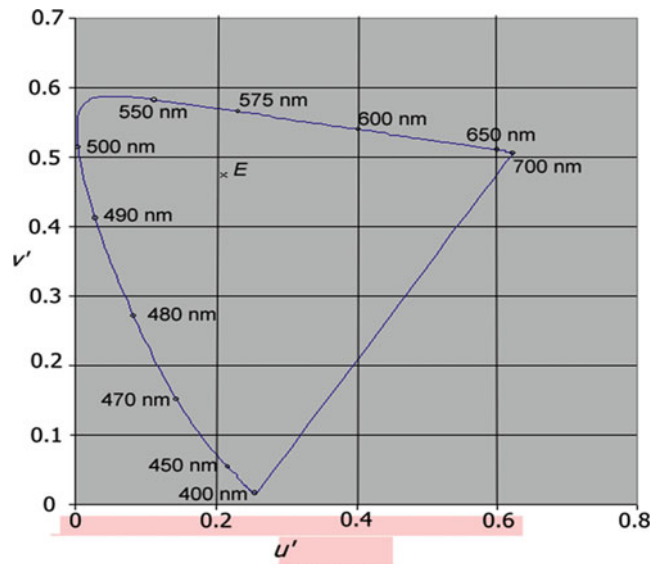
$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z} \end{aligned} \quad (3)$$

To be exact Euclidean distances in his diagram can be used to represent approximately the relative perceived magnitude of color differences between color stimuli of negligibly different luminances, of approximately the same size, and viewed in identical surroundings, by an observer photopically adapted to a field with the chromaticity of CIE standard illuminant D65 [6].

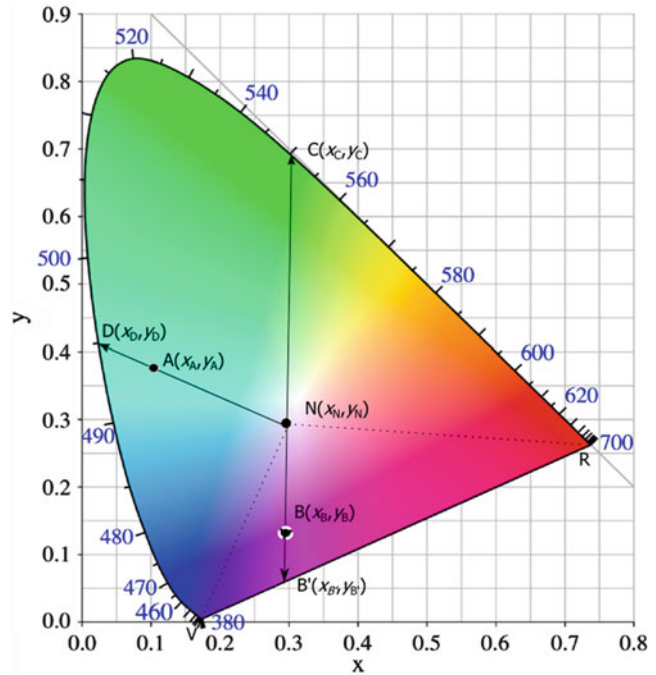
Dominant Wavelength and Purity

A color can be characterized by its tristimulus values or its chromaticity and the luminance (if it is a self-luminous object) or luminance factor (if it is a reflecting or transmitting object illuminated by a (standard) light source). It is difficult to visualize the chromaticity from the x , y values; an easier

CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength,
Fig. 2 CIE u' , v' chromaticity diagram



CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength, Fig. 3 A colored chromaticity diagram, where the basic components of excitation purity are shown



identification is by two other quantities: dominant wavelength and excitation purity.

Dominant and Complementary Wavelength

In Fig. 3, we see two colored samples (represented in the chromaticity diagram by A and B); they are illuminated by a source of neutral chromaticity (N). If a line is drawn from point N through point A or B, one reaches at the boundaries of the chromaticity diagram, at the spectrum locus points D and C, respectively. Points D and A are located on the same side of point N; thus, chromaticity of A is less saturated as that of D but has similar hue; therefore, the wavelength of the monochromatic radiation at point D is called the *dominant wavelength* (in our example 495 nm). By mixing radiation of the monochromatic radiation D and the neutral radiation N, one can create the chromaticity A.

For point B, as it is located on the far side of points N and C, one can produce the chromaticity N by mixing chromaticity B with C. Therefore, the wavelength of the spectral line at C is called the *complementary wavelength* for chromaticity B.

Excitation Purity

The relative distance of A (resp. B) from N, compared to the distances $\overline{DN}$ (resp. $\overline{B'N}$), is called *excitation purity* and describes how strongly the monochromatic stimulus is diluted by the radiation of N. For purple colors (in the triangle of points N-V-R), the monochromatic stimulus is replaced, as seen, by the stimulus on the purple boundary. In practice it is not necessary to calculate with the vector length, it is enough to take either the x or the y coordinates. One should take always those coordinates that are larger; thus, e.g., for the two chromaticity points A and B, the excitation purities are calculated as

Excitation purity of chromaticity point A :

$$p_{e,A} = \frac{x_A - x_N}{x_D - x_N} \quad (4)$$

Excitation purity of chromaticity point B :

$$p_{e,B} = \frac{y_B - y_N}{y_{B'} - y_N} \quad (5)$$

Colorimetric Purity

As mentioned under Definitions, another purity quantity, *colorimetric purity*, is defined by the luminance of the respective stimuli: Given the stimulus A, to mix this color from stimuli D and N, one needs luminance L_D and L_N . With these quantities, the colorimetric purity is

$$p_c = \frac{L_D}{L_N + L_D}. \quad (6)$$

Summary

With some practice, one gets a reasonable feeling of the chromaticity of monochromatic stimuli if their wavelength is given; thus, if the dominant/complementary wavelength of a stimulus is stated, one can form a mental picture of the stimulus. Similarly the excitation purity is also a relatively easily visualized quantity – how whitish the given colored stimulus is – thus, these two quantities are often used instead of the chromaticity coordinates for a quick description of the chromaticity of a stimulus. One has to emphasize – however – that the chromaticity of the neutral stimulus is important. In many colorimetric calculations, the CIE standard illuminant D65 is used as a reference neutral stimulus, but in some applications, the equienergetic stimulus is found.

The CIE 1931 x, y chromaticity diagram is the most often used diagram. It is, however, non-equidistant, i.e., in different parts of the chromaticity diagram, perceived equal chromaticity differences are observed as different coordinate differences. The CIE 1976 u', v' diagram is more equidistant and is generally used in lighting engineering.

There is one exception, the determination of correlated color temperature, which is determined in the CIE 1960 diagram, the coordinates of which are the following:

$$u = u', \quad v = \frac{2}{3} v'. \quad (7)$$

For further details, see Eq. 2.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE \$u', v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

References

1. Commission Internationale d'Eclairage: International lighting vocabulary. CIE S 017/E:2011. see also <http://eilm.cie.co.at>
2. Commission Internationale d'Eclairage: Colorimetry – Part 3: CIE Tristimulus Values. CIE S 14-3/E (2011)
3. Commission Internationale d'Eclairage: Colorimetry, 3rd edn. CIE 015 (2004)
4. Schanda, J.: CIE colorimetry, Chap 3. In: Schanda, J. (ed.) CIE Colorimetry – Understanding the CIE System. Wiley Interscience (2007)
5. Commission Internationale d'Eclairage: Colorimetry – Part 5: CIE 1976 $L^*u^*v^*$ Colour Space and u', v' Uniform Chromaticity Scale Diagram. CIE S 014-5/E (2009)
6. Commission Internationale d'Eclairage: Colorimetry – Part 2: CIE Standard Illuminants. CIE S 124-2/E (2007)

CIE Color Appearance Model 2002

- [CIECAM02](#)

CIE Color Rendering Index

Hirohisa Yaguchi
Professor Emeritus, Chiba University, Chiba,
Japan

Synonyms

[Color fidelity index](#)

Definition

Color rendering of a light source is the effect of an illuminant on the perceived color of objects by conscious or subconscious comparison with their perceived color under a reference illuminant (CIE S 017:2020 ILV: International Lighting Vocabulary, 2nd edition).

In this content, it is appropriate to define a color rendering index of a light source as a measure of the degree to which the perceived color of an object illuminated by the test illuminant conforms to that of the same object illuminated by the reference illuminant, for specified conditions.

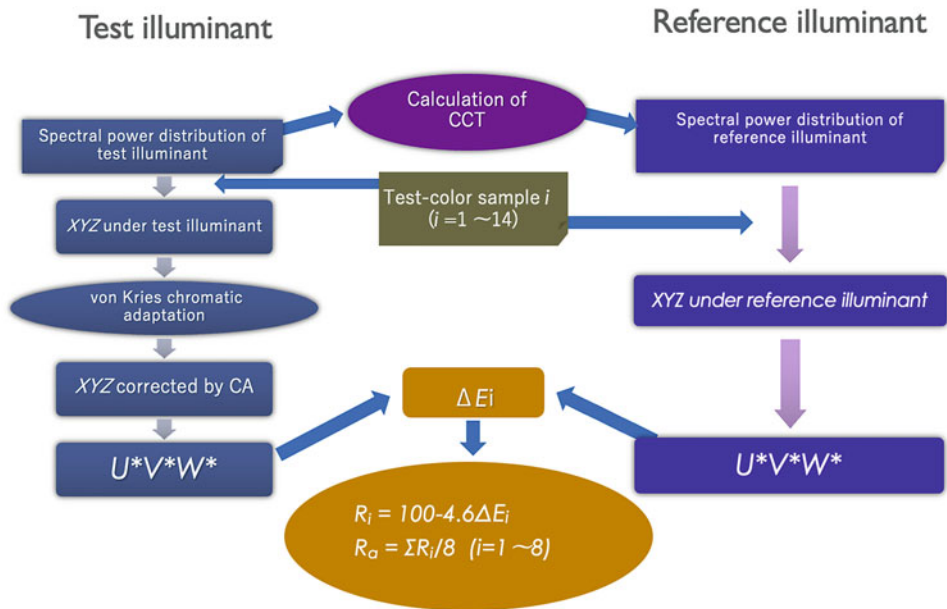
Overview

The color rendering index of the light source is calculated on the basis of the color difference between the tristimulus values of the test-color samples illuminated with the test light source and the tristimulus values of the same test-color sample illuminated with the reference light source having the same correlated color temperature (CCT) as the test light source. For the reference

illuminant, a Planckian radiator is used at low temperatures of 5000 K or less, and CIE daylight illuminant is used at higher than 5000 K, depending on the CCT of the test light source. The special color rendering index R_i is scaled so that $R_i = 100$ is obtained when the tristimulus values of test-color sample i under the test light source are the same as those calculated with respect to the reference illuminant. The general color rendering index R_a is calculated from the arithmetic mean of the special rendering indices for the eight test-color samples $i = 1$ to 8.

Determination of Color Rendering Indices

Figure 1 shows the flow chart to the determination of the CIE color rendering indices (1995). The first step is to select a reference light source. In order to determine a reference illuminant that has the same CCT as the test illuminant, start by calculating the CCT of the test illuminant. Next, the tristimulus values of a test-color sample illuminated with a test light source and those of a reference light source with the same CCT are



CIE Color Rendering Index, Fig. 1 The flow to the determination of the CIE color rendering index

calculated. As the test light source might be deviated from the chromaticity of the reference illuminant along the line of constant CCT, one should allow for chromatic adaptation by means of a von Kries of transformation with reference stimuli. The final step is to calculate the color difference from the tristimulus values of a test-color sample under the test illuminant and those under the reference illuminant. The smaller the color difference, the closer the appearance of the object color under the test light source to that under the reference illuminant. So, the smaller color difference gives the better color rendition. The final formula of the CIE color rendering indices have been calculated based on the color difference between the tristimulus values of a test-color sample under the test light source illuminant and those under the reference illuminant.

Reference Illuminant

The choice of a reference light source compared to the test light source should be governed by whatever is considered to be the original object color perception; that is, color appearance of an object in the light in which it is usually observed. In most practical cases, this light will be incandescent lamp or natural daylight. So, CIE determined that the reference illuminant for light sources with a CCT below 5000 K shall be a Planckian radiator and from 5000 K, one of a series of spectral radiant power distributions of phases of CIE daylight. The reference illuminant is intended to be of the same or nearly the same chromaticity as the test light source. The acceptable chromaticity difference between the test source and the reference illuminant corresponds to 15 MK⁻¹ along the Planckian locus in the chromaticity diagram.

Test-Color Samples

For calculating the CIE general color rendering indices, a set of eight test-color samples is specified in terms of the spectral reflectance. For

CIE Color Rendering Index, Table 1 Munsell notations of the test-color samples

Sample number	Munsell notation	Sample number	Munsell notation
1	7.5R6/4	9	4.5R4/13
2	5Y6/5	10	5Y8/10
3	5GY6/8	11	4.5G5/8
4	2.5G6/6	12	3 PB3/11
5	10BG6/4	13	5YR8/4
6	5 PB6/8	14	5GY4/4
7	2.5P6/8		
8	10P6/8		

calculating the CIE special color rendering indices, any one of the six test-color samples in addition to any one from the above set of eight test-color samples may be used. Their approximate Munsell notations are listed in Table 1.

Color Difference

The color difference between the tristimulus values of the test-color samples illuminated with the test source and the reference illuminant is calculated using the CIE 1964 uniform color space, so called CIE U*V*W* space. The CIE 1964 uniform color space is now outdated and replaces by the CIE 1976 uniform color space, but the color rendering calculations are still performed in the CIE U*V*W* space. The color difference between a test-color sample numbered *i* illuminated by the test and the reference illuminant is denoted as Δ*E_i*.

CIE Color Rendering Indices

A CIE special color rendering index *R_i* is calculated on the basis of the color difference Δ*E_i* for *s* test-color samples *i*, by means of the following formula:

$$R_i = 100 - 4.6\Delta E_i$$

The index *R_i* has been scaled so that *R_i* = 100 is obtained when the tristimulus values of the

test-color sample i under the test source are the same as those calculated with respect to the reference illuminant. So, $R_i = 100$ means the perfect rendition for the test-color sample. The constant factor 4.6 has been selected to give a value of $R_i = 50$ for a traditional warm white fluorescent lamp.

The CIE general color rendering index is defined as the arithmetic mean of the eight test-color samples $i = 1$ to 8:

$$R_a = \frac{1}{8} \sum_{i=1}^8 R_i$$

References

1. CIE: CIE 13.3 method of measuring and specifying colour rendering properties of light sources. CIE, Vienna (1995)
2. CIE S 017:2020 ILV: International Lighting Vocabulary, 2nd edition

CIE Color-Matching Functions

► [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)

CIE Cone Fundamentals

► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)

CIE Fundamental Color Matching Functions

► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
 ► [Cone Fundamentals](#)

CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156)

Jan Morovic

Hewlett-Packard Company, Sant Cugat del Valles/Barcelona, Catalonia, Spain

Definition

The *CIE Guidelines for the Evaluation of Gamut Mapping Algorithms* (referred to as *Guidelines* in the remainder of this entry) set out experimental conditions under which color gamut mapping algorithms are to be evaluated so that results can be compared and combined from separate experiments. The *Guidelines* were published [1] in 2004 by Division 8 of the CIE and cover a number of aspects of experimental evaluation, both mandatory and optional. They also include case studies for applying them to various color reproduction scenarios and a checklist that can be used to determine an experiment's compliance with the *Guidelines*.

Overview

A color gamut mapping algorithm is that part of a color reproduction process, which ensures that colors from some original (source) are adapted to fit inside the color gamut available under reproduction (destination) conditions. A typical example is a color image viewed on an electronic display that is to be reproduced in print. Here, there are colors that a display can generate (e.g., bright greens), which cannot be matched in print, and a substitution of the out-of-gamut color by an in-gamut color needs to be made. Note that the converse – representing printed colors on a display – also requires gamut mapping, since some printable colors (e.g., cyans) are beyond the capabilities of displays, and that this is the case for the great majority of original–reproduction combinations.

Since gamut mapping can have different aims (e.g., resulting in most similar reproduction to the original or resulting in a most pleasing reproduction), since its performance cannot be measured, and since optimal performance cannot be known in absolute terms, determining how well a gamut mapping algorithm works or whether one algorithm outperforms another is a challenge. Reviewing the literature on this subject [2] prior to the *Guidelines*' publication reveals a great variety of experimental methods and conditions used for comparing alternative gamut mapping algorithms that yielded incomparable but seemingly contradictory conclusions.

The *Guidelines* therefore set out recommendations about the following aspects of a gamut mapping algorithm's evaluation: test images (and their exchange), original and reproduction media, viewing conditions, color measurement, gamut boundary description, a pair of reference gamut mapping algorithms, the color spaces in which they are to be applied, and the psychovisual experimental method to be used. In each of these sections, there are both obligatory aspects and recommended ones, and attention is also paid to how experimental details are to be reported. Finally, the *Guidelines* also show how they are to be applied for specific, typical workflows (Reference Output Medium Metric (ROMM) RGB to print, Cathode-Ray Tube (CRT) to print, and transparency to print), an appendix elaborates on some of the *Guidelines*' sections, and a checklist is provided to determine an experiment's compliance with the *Guidelines*.

Before proceeding to the specifics of experimental evaluation, the *Guidelines* specify the obligatory use of the "subjective accuracy" color reproduction intent, which is defined as aiming at "reproducing a given colour image in a way where the reproduction is as close to the original as possible, this similarity is determined psychophysically and the process has no image enhancing aims." Evaluation of other reproduction intents is not discouraged, but subjective accuracy is mandatory.

Test images are specified first, with the one mandatory image, "Ski" (Fig. 1), made available as a physical transparency, in ROMM RGB and

sRGB (i.e., two colorimetrically defined RGB color spaces) rendered color encodings, and as CIELAB version of these three cases. The *Guidelines* further specify the need to share with the CIE any other test images used in compliant experiments and detail how such sharing is to be done.

Reproductions of the obligatory and other test images then need to be made on a combination of color reproduction media from among the following four types: reflective print, transparency, monitor, and virtual (i.e., wide gamut color spaces such as scRGB [3] or ROMM RGB [4]). Constraints are imposed on acceptable uniformity, repeatability, and viewing geometry-dependent characteristics of these media, and the *Guidelines* also specify which of the media's aspects to report (including measured characterization data).

Given a set of chosen test images, rendered on one medium as originals and to be gamut mapped to another medium, the viewing conditions under which the media are to be viewed are specified next. Here, it is mandatory to report "chromaticity and luminance level of the white point, level and correlated colour temperature of the ambient illumination, light source colour rendering index and nature of field of view (proximal field, background, surround)." A recommendation is made about the original and reproduction images being the same size, the illumination having a color rendering index of at least 90, and the uniformity of illumination dropping off to no less than 75% of central peak illumination. Control over the viewing environment needs to be exercised to exclude extraneous light sources as well as light reflected from objects in it, and specific border and surround characteristics are mandated. Specific luminance levels for different types of media, following ISO 3664 [5], also need to be ensured, and particular care is taken to define aspects of monitor to print matching, where three alternatives are offered for how the two media's white points are to relate: an absolute colorimetric match at D65 chromaticity, an adherence to the per media specifications (i.e., D65 for monitor and D50 for print), or a D65 chromaticity match but at luminance levels as mandated per media.

In terms of color measurement, the requirement is to carry it out as closely to experimental

Specifically, SGCK involves the following transformation for each original color:

1. Keep h_{ab}^* unchanged.
2. Map lightness as follows:

$$L_r^* = (1 - p_C) L_o^* + p_C L_s^* \quad (1)$$

where o refers to the original, r refers to the reproduction, and p_C is a chroma-dependent weight computed from the original color's C^* :

$$p_C = 1 - ((C^{*3}) / (C^{*3} + 5 \times 10^5))^{1/2} \quad (2)$$

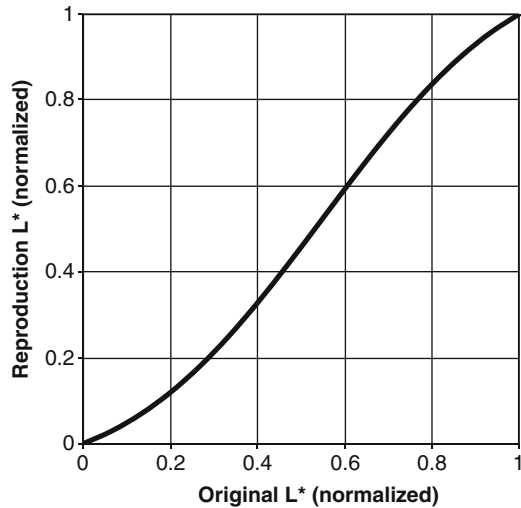
and where L_s^* is the result of the original's L_o^* being mapped using the following sigmoidal function (Fig. 2), having x_0 and Σ parameters empirically derived for different levels of reproduction medium minimum L^* (e.g., for a minimum reproduction L^* of 15, $x_0 = 58.2$ and $\Sigma = 35$):

$$S_i = \sum_{n=0}^{n=i} \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{(\frac{100n}{\Sigma} - x_0)^2}{2\Sigma^2}} \quad (3)$$

S_i values generate using Function 3 for $i \in [0, m]$ and then form a look-up table that is further scaled using the L^* ranges of the original and reproduction:

$$S_{LUT} = \frac{(S_i - \min(S))}{(\max(S) - \min(S))} (L_{\max r}^* - L_{\min r}^*) + L_{\min r}^* \quad (4)$$

Finally, the L_s^* value needed for Eq. 1 can be



CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156), Fig. 2 Sigmoidal function used for L^* mapping (for $x_0 = 58.2$ and $\Sigma = 35$)

obtained by interpolating in the S_{LUT} look-up table with an L_o^* modified as follows:

$$L_{o'}^* = 100 (L_o^* - L_{\min o}^*) / (L_{\max o}^* - L_{\min o}^*) \quad (5)$$

3. The original's C^* and L_r^* obtained from Eq. 1 are next mapped in a plane of constant h_{ab}^* toward the L^* of the cusp (the color with maximum C^* in the reproduction gamut at this h_{ab}^*) as follows:

$$d_r = \begin{cases} d_o; d_o \leq 0.9d_{gr} \\ 0.9d_{gr} + (d_o - 0.9d_{gr})0.1d_{gr}/(d_{go} - 0.9d_{gr}); d_o > 0.9 \times d_{gr} \end{cases} \quad (6)$$

where g refers to gamut boundary, o and r to original and reproduction, and d to distance from the cusp's L^* on the L^* axis (Fig. 3).

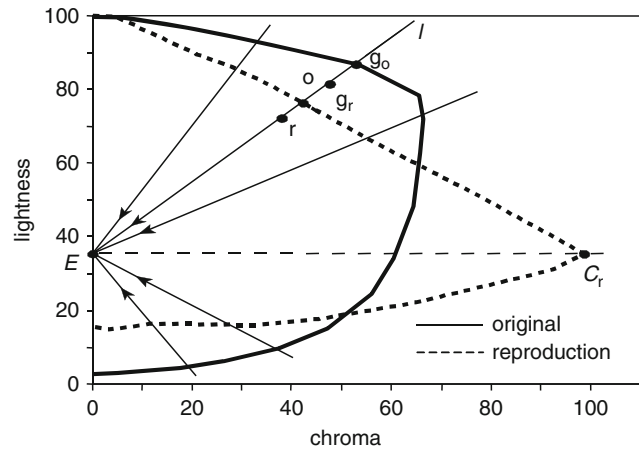
Gamut mapping algorithms that are compared among themselves and versus HPMINDE and

SGCK need to be described with sufficient detail for repeatability.

In terms of color spaces in which gamut mapping is to be performed, the *Guidelines* again only require the reporting of whichever space was used. A recommendation is made for isotropic

**CIE Guidelines for
Evaluation of Gamut
Mapping Algorithms:
Summary and Related
Work (Pub. 156),**

Fig. 3 L^* and C^* mapping
toward the L^* of the cusp



color spaces that have greater hue uniformity than CIELAB (e.g., IPT [9], ► [CIECAM02](#)).

Finally, the pair comparison, category judgment, and ranking methods are proposed as alternatives for how reproductions made using different gamut mapping algorithms are to be compared visually. Apart from the obligation to use at least 15 observers, there is an extensive list of experimental aspects that need to be reported, and Appendix C provides background on them.

To aid the application of the *Guidelines*, three common scenarios are described in more detail, and recommendations are made for what choices to make in terms of the *Guidelines*' parts. The scenarios are ROMM to print, CRT to print, and transparency to print, and Appendix D includes a checklist that can be completed to test compliance with the *Guidelines*.

Future Directions

The *Guidelines* have been used extensively since their publication in 2004 to inform the design and execution of the experimental evaluation of gamut mapping algorithms. A notable aspect of these experiments is the use of greater numbers of test images, such as 15 [10], 20 [11], and even 250 [12], as compared to the previous trend of using around five, which has contributed to a general unreliability of results. The *Guidelines* have also been used by a large-scale evaluation

of nine printer manufacturers' products reported by Fukasawa et al. [13]. Since they were formulated close to 10 years ago, there are aspects of the *Guidelines*, e.g., their reference to CRTs and film transparencies and lack of reference to wide gamut displays, which would benefit from future revision. The *Guidelines* also prepared the ground for interrelating the results of multiple, compliant experiments, which too is yet to be implemented.

References

1. CIE Publication 156: Guidelines for the evaluation of gamut mapping algorithms. (2004)
2. Morovic, J.: Color Gamut Mapping. Wiley Chichester, UK (2008)
3. IEC 61966-2-2: Multimedia systems and equipment – colour measurement and management – part 2-2: colour management – extended RGB colour space – scRGB (2003)
4. Spaulding, K.E., Woolfe, G.J., Giorgianni, E.J.: Reference input/output medium metric RGB color encodings (RIMM/ROMM RGB). In: IS&T PICS 2000 Conference Proceedings, pp. 155–163, Portland (2000)
5. ISO 3664: Viewing conditions – prints, transparencies and substrates for graphic arts technology and photography (2000)
6. Morovic, J., Sun, P.L.: Non-iterative minimum ΔE gamut clipping. In: IS&T/SID 9th Color Imaging Conference, pp. 251–256, Scottsdale (2001)
7. Morovic, J.: To Develop a Universal Gamut Mapping Algorithm. Ph.D. thesis, University of Derby (1998)
8. Braun, G.J., Fairchild, M.D.: Image lightness rescaling using sigmoidal contrast enhancement functions. *J. Electron. Imaging*. **8**(4), 380–393 (1999)

9. Ebner, F., Fairchild, M.D.: Development and testing of a color space (IPT) with improved hue uniformity. In: Proceedings of 6th IS&T/SID Color Imaging Conference, pp. 8–13. Scottsdale (1998)
10. Bonnier, N., Schmitt, F., Brettel, H., Berche, S.: Evaluation of spatial gamut mapping algorithms. In: IS&T/SID 14th Color Imaging Conference, pp. 56–61, Scottsdale (2006)
11. Dugay, F., Farup, I., Hardeberg, J.Y.: Perceptual evaluation of color gamut mapping algorithms. *Color. Res. Appl.* **33**(6), 470–476 (2008)
12. Zolliker, P., Dätwyler, M., Simon, K.: Gamut mapping for small destination gamuts. In: AIC Colour 05–10th Congress of the International Colour Association, pp. 345–348, Granada, Spain (2005)
13. Fukasawa, K., Ito, A., Qunigoh, M., Nakaya, F., Shibuya, T., Shimada, H., Yaguchi, H.: Suitable printer color reproduction for office environment. In: Proceedings of the 13th IS&T/SID Color Imaging Conference, pp. 185–188. Scottsdale (2005)

CIE Guidelines for Mixed Mode Illumination: Summary and Related Work

Suchitra Sueeprasan

Department of Imaging and Printing Technology,
Chulalongkorn University Intellectual
Repository, Chulalongkorn University, Bangkok,
Thailand

Synonyms

Mixed adaptation condition; Mixed chromatic adaptation

Definition

Mixed mode illumination refers to an image comparison between softcopy and hardcopy with successive binocular viewing. When comparing the softcopy images on self-luminous displays with the hardcopy images under ambient lighting, an observer's eyes move back and forth between the images. Under such circumstance, the state of adaptation is unfixed and the human visual system partially adapts to the white point of the softcopy

display and partially adapts to the ambient illumination. The term *mixed chromatic adaptation* is defined as a state in which observers adapt to light from sources of different chromaticities [3].

Background

In 1998, the Technical Committee 8–04, Adaptation under Mixed Illumination Conditions, was formed in Commission Internationale de l'Eclairage (CIE)/Division 8 (Image Technology), with the aim to investigate the state of adaptation of the visual system when comparing softcopy images on self-luminous displays and hardcopy images viewed under various ambient lighting conditions. A number of experiments were conducted with the goal to achieve color appearance matches between softcopy and hardcopy images under mixed illumination conditions. Katoh [5–7] developed the mixed adaptation model, namely, S-LMS, for such application. In his study, softcopy images on a CRT were compared with hardcopy images under F6 illumination. Under equal luminance levels of softcopy and hardcopy, it was found that the human visual system was 60% adapted to the monitor's white point and 40% to the ambient light. The same adaptation ratio was also found for unequal luminance levels. It is concluded that the adaptation ratio was independent of image content, luminance, and chromaticity of the monitor's white point and the ambient illumination.

The study by Berns and Choh [1], in which softcopy images were compared with hardcopy images under F2 illuminant with equal luminance levels, showed that an image with a chromatic adaptation shift of 50% was most preferred as the closest color match and the best stand-alone image. The color model tested in this study was the RLAB color space. Shiraiwa et al. [9] proposed a new method in which the mixed chromatic adaptation was applied in CIE xy chromaticity coordinates. The best adaptation ratio was between 50% and 60%, which is similar to the previous studies. In the visual experiments, where the illuminants of softcopy and hardcopy images were different, their proposed method and

the S-LMS model, both incorporating the mixed adaption, generated better color appearance matches than the conventional color management systems.

Henley and Fairchild [4] tested the performance of various color models with the inclusion and exclusion of mixed adaptation. Observers made appearance matches of color patches on a CRT to hardcopy originals under six different matching methods. The results were reported in terms of color differences between the actual match and the predicted match by color models. The models incorporating the mixed adaptation improved the results in all conditions over their corresponding conventional models.

The studies by Katoh and Nakabayashi [8] and Sueeprasan and Luo [10] closely followed the experimental guidelines put forth by TC8–04. In the experiments, softcopy and hardcopy images were compared using the simultaneous binocular matching technique. Various color models were tested. In Katoh and Nakabayashi's study, the linear transformation matrix in the S-LMS model was replaced by different chromatic adaptation transform matrices. The results showed that the S-LMS model with the Bradford (BFD) matrix performed best. They also investigated whether incomplete adaptation was needed in the mixed chromatic adaptation model. RLAB method and D factor resulted in much better score than complete adaptation, indicating that the incomplete adaptation was essential.

In Sueeprasan and Luo's study, the performance of the promising chromatic adaptation transforms (CMCCAT97, CMCCAT2000, and CIECAT94) and the S-LMS mixed chromatic adaptation transform was compared. The state of chromatic adaptation was also investigated. The results showed that the incomplete adaptation ratio was crucial in producing color matches. The human visual system was between 40% and 60% adapted to the white point of the monitor regardless of the changes in illumination conditions. CMCCAT2000 outperformed the other models.

The results from the previous studies are in good agreement for the chromatic adaptation ratio, which is in the range of 40–60% adapted

to the white point of the monitor. The adaptation ratio is consistent over various viewing conditions and regardless of the chromatic adaptation transform and incomplete adaptation formula used. Both incomplete and mixed chromatic adaptations are required in the mixed adaptation model for predicting color matches under mixed illumination conditions. Based on these findings, TC8–04 recommends the mixed adaptation model for use in cross-media color reproduction when mixed illumination conditions are employed.

Recommended Model

CIE TC8–04 recommends the S-LMS mixed adaptation model for achieving appearance matches under mixed chromatic adaptation. The S-LMS model is fundamentally a modified form of the von Kries transformation with incorporation of partial adaptation. The compensation for chromatic adaptation includes incomplete adaptation and mixed adaptation.

The first step of the S-LMS model is to transform XYZ tristimulus values to the cone signals for the human visual system (Eqs. 1, 2, and 3). $X_n Y_n Z_n$ values are tristimulus values of the reference white. M_{CAT02} is the chromatic adaptation transformation matrix used in CIECAM02 [2].

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = M_{CAT02} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} L_{n(CRT)} \\ M_{n(CRT)} \\ S_{n(CRT)} \end{bmatrix} = M_{CAT02} \begin{bmatrix} X_{n(CRT)} \\ Y_{n(CRT)} \\ Z_{n(CRT)} \end{bmatrix} \quad (2)$$

$$M_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6974 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (3)$$

Then, compensation is made for the change in chromatic adaptation according to the

surroundings. The human visual system changes the cone sensitivity of each channel to compensate for the change in illumination. In the calculation, the signals of each channel are divided by those of the adapted white. There are two steps of calculation to obtain the adapted white of a monitor.

The first step is the compensation for the incomplete adaptation of the visual system to the self-luminous displays (Eq. 4). Even if the monitor is placed in a totally dark room, the chromatic adaptation of the human visual system to the white point of the monitor will not be complete. That is, the reference white of the monitor does not appear perfectly white. Chromatic adaptation becomes less complete as the chromaticity of the adapting stimulus deviates from the illuminant E and as the luminance of the adapting stimulus decreases. The D factor from CIECAM02 is used (Eq. 5). F is 1.0, and L_A is the absolute luminance of the adapting field.

$$\begin{aligned} L'_{n(\text{CRT})} &= L_{n(\text{CRT})} / \{D + L_{n(\text{CRT})}(1 - D)\} \\ M'_{n(\text{CRT})} &= M_{n(\text{CRT})} / \{D + M_{n(\text{CRT})}(1 - D)\} \\ S'_{n(\text{CRT})} &= S_{n(\text{CRT})} / \{D + S_{n(\text{CRT})}(1 - D)\} \end{aligned} \quad (4)$$

$$D = F \left\{ 1 - \left(\frac{1}{3.6} \right) e^{\left(\frac{L_A - 42}{92} \right)} \right\} \quad (5)$$

The next step is the compensation for mixed adaptation. In cases where the white points of the monitor and the ambient light are different, it was hypothesized that the human visual system is partially adapted to the white point of the monitor and partly to the white point of the ambient light. Therefore, the adapting stimulus for softcopy images can be expressed as the intermediate point of the two (Eqs. 6 and 7). It should be noted that incompletely adapted white is used for the white point of the monitor. $Y_{n(\text{CRT})}$ is the absolute luminance of the white point of the monitor, and Y_{ambient} is the absolute luminance of the ambient light.

$$\begin{aligned} L''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left(\frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot L'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left(\frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot L_{\text{ambient}} \\ M''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left(\frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot M'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left(\frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot M_{\text{ambient}} \\ S''_{n(\text{CRT})} &= R_{\text{adp}} \cdot \left(\frac{Y_{n(\text{CRT})}}{Y_{\text{adp}}} \right)^{1/3} \cdot S'_{n(\text{CRT})} \\ &\quad + (1 - R_{\text{adp}}) \cdot \left(\frac{Y_{\text{ambient}}}{Y_{\text{adp}}} \right)^{1/3} \cdot S_{\text{ambient}} \end{aligned} \quad (6)$$

$$Y_{\text{adp}} = \left\{ R_{\text{adp}} \cdot Y_{n(\text{CRT})}^{1/3} + (1 - R_{\text{adp}}) \cdot Y_{\text{ambient}}^{1/3} \right\}^3 \quad (7)$$

When the luminance of the monitor, $Y_{n(\text{CRT})}$, equals the ambient luminance, Y_{ambient} , the adapting white can be calculated by Eq. 8.

$$\begin{aligned} L''_{n(\text{CRT})} &= R_{\text{adp}} \cdot L'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot L_{\text{ambient}} \\ M''_{n(\text{CRT})} &= R_{\text{adp}} \cdot M'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot M_{\text{ambient}} \\ S''_{n(\text{CRT})} &= R_{\text{adp}} \cdot S'_{n(\text{CRT})} + (1 - R_{\text{adp}}) \cdot S_{\text{ambient}} \end{aligned} \quad (8)$$

R_{adp} is the adaptation ratio to the white point of the monitor. When the ratio R_{adp} equals 1.0, the human visual system is assumed to be fully adapted to the white point of the monitor and none to the ambient light. Conversely, when the ratio is 0.0, the human visual system is assumed to be completely adapted to the ambient light and none to the white of the monitor. These two extreme cases assume that the human visual system is at single-state chromatic adaptation. For mixed chromatic adaptation, 0.6 is chosen for R_{adp} in the S-LMS model.

With the newly defined white points for the softcopy images, the von Kries chromatic adaptation model is applied. The cone signals after adaptation are calculated as Eq. 9.

$$\begin{aligned} L_s &= L_{(\text{CRT})}/L''_{n(\text{CRT})} \\ M_s &= M_{(\text{CRT})}/M''_{n(\text{CRT})} \\ S_s &= S_{(\text{CRT})}/S''_{n(\text{CRT})} \end{aligned} \quad (9)$$

For hardcopy images, the simple von Kries chromatic adaptation without incomplete chromatic adaptation and mixed chromatic adaptation is used (Eq. 10). The paperwhite is chosen as the reference white, because the eye tends to adapt according to the perceived whitest point of the scene.

$$\begin{aligned} L_s &= L_{(\text{Print})}/L_{n(\text{Print})} \\ M_s &= M_{(\text{Print})}/M_{n(\text{Print})} \\ S_s &= S_{(\text{Print})}/S_{n(\text{Print})} \end{aligned} \quad (10)$$

Implementation

The S-LMS mixed adaptation model is designed to integrate into the CIECAM02 model, which is developed for color management applications. Hence, the S-LMS model should be applied to extend the CIECAM02 model for use in cross-media color reproduction when mixed mode illumination is employed. The input parameters required are XYZ tristimulus values of the white point of the monitor, the adopted white for the ambient light, and the paper white of hardcopy.

Cross-References

- [Adaptation](#)
- [CIECAM02](#)

References

1. Berns, R.S., Choh, K.H.: Cathode-ray-tube to reflection-print matching under mixed chromatic adaptation using RLAB. *J. Electron. Imaging*. **4**, 347–359 (1995)
2. CIE Publication 159: A colour appearance model for colour management systems: CIECAM02. (2004)
3. CIE Publication 162: Chromatic adaptation under mixed illumination condition when comparing softcopy and hardcopy images. (2004)
4. Henley, S.A., Fairchild, M.D.: Quantifying mixed adaptation in cross-media color reproduction. In: Proceedings of the 8th IS&T/SID Color Imaging Conference, pp. 305–310 (2000)
5. Katoh, N.: Practical method for appearance match between soft copy and hard copy. *SPIE Proc.* **2170**, 170–181 (1994)
6. Katoh, N.: Appearance match between soft copy and hard copy under mixed chromatic adaptation. In: Proceedings of IS&T 3rd Color Imaging Conference, pp. 22–25 (1995)
7. Katoh, N., Nakabayashi, K.: Effect of ambient light on color appearance of soft copy images—color management on the network. In: Proceedings of AIC Color 97 Kyoto 2, pp. 582–585. (1997)
8. Katoh, N., Nakabayashi, K.: Applying mixed adaptation to various chromatic adaptation transformation (CAT) models. In: PICS Conference Proceedings, pp. 299–305. (2001)
9. Shiraiwa, Y., Hidaka, Y., Mizuno, T., Sasaki, T., Ohta, K., Usami, A.: Matching of the appearance of colors in hard-copy and soft-copy images in different office environments. *SPIE Proc.* **3300**, 148–158 (1998)
10. Sueeprasan, S., Luo, M.R.: Applying chromatic adaptation transforms to mixed adaptation conditions. *Color. Res. Appl.* **28**(6), 436–444 (2003)

CIE L*a*b*

- [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

CIE Method of Assessing Daylight Simulators

Robert Hirschler
Colour Consultant, Budapest, Hungary

Definition

Daylight simulator is a “device that provides spectral irradiance approximating that of a CIE standard daylight illuminant or CIE daylight illuminant, for visual appraisal or measurement of colours” [1].

Development of the CIE Methods of Assessing Daylight Simulators

Although for decades after the acceptance of the CIE system (1931), Illuminant C was accepted and widely used in colorimetry, the practical implementation of Source C was limited to special laboratory use. In 1963, the Colorimetry Committee of the CIE decided to supplement the then existing CIE illuminants A, B, and C by new illuminants more adequately representing phases of natural daylight. These new illuminants (D55, D65, and D75) were defined by a new approach suggested by Judd et al. [2] based on Simonds' [3] method of reducing experimental data to characteristic vectors (eigenvectors) and calculating the relative spectral power distribution of daylight of any desired correlated color temperature.

As a consequence of this new approach, the new daylight illuminants have been defined theoretically (albeit based on experimental data), and it means that there are still no physically realizable light sources corresponding to the illuminants. It was clear to the CIE already in 1967 that daylight simulators were required that would serve as standard sources representing the daylight illuminants. As a first step, Wyszecki [4] published spectral irradiance distribution data on a number of daylight simulators and also suggested methods for evaluating how well these sources simulated the corresponding illuminants.

When asking the question "how close" a given source is to the illuminant of the same correlated color temperature, we must first decide how to measure this closeness.

The "fingerprints" of illuminants and light sources are their spectral power distributions, and for most of the colorimetric calculations, only the relative values are interesting, calculated from the spectral radiance or irradiance values and normalized to have the value of 100 at 560 nm or to have $Y = 100$.

The colorimetric properties of illuminants and light sources are generally described in terms of the x, y or u', v' chromaticity coordinates, the correlated color temperature, and very often the color rendering index. Although these properties may often give us sufficient information, for the more

specific purpose of evaluating whether a given light source may or may not be considered an adequate realization (simulation) of the corresponding illuminant, more complex measures are needed. By and large, they can be divided into two groups: those comparing the spectral curves with or without applying weights to take the visual significance into consideration and those measuring the effect of the illumination on a selected group of object colors and then comparing either the change from one illuminant to the other (color rendering index or CRI type) or by calculating the color difference under the test source illuminant for pairs of samples which are perfect matches under the reference illuminant (metamerism index or MI type). In the early 1970s, a subcommittee of TC-1.3 on Standard Sources studied a number of proposals for different methods: based on MI-type indices for the visible range [5–7] and based on the effective excitation of three fluorescent samples for the UV range [8]. At the Troy (1977) meeting of TC-1.3, a modification of the Ganz [9] proposal was adopted for the UV range evaluation: the method used three virtual metameric (isomeric) pairs for each illuminant, each consisting of a fluorescent and a nonfluorescent sample.

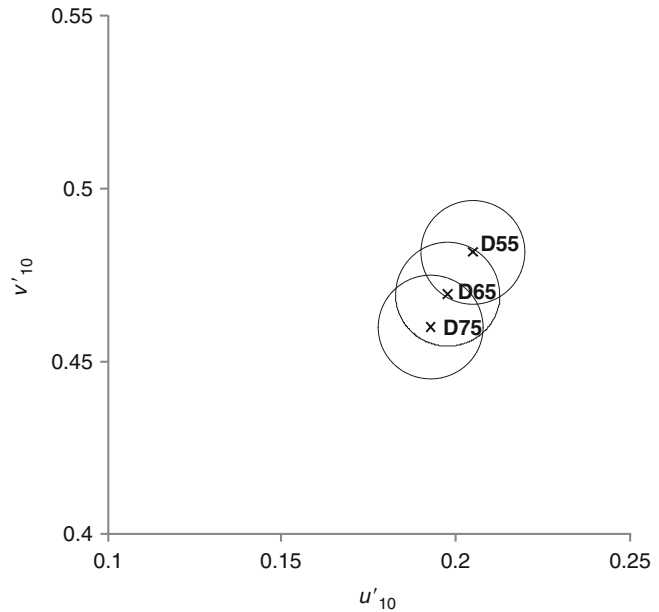
The recommendations of TC-1.3 were first published in 1981 as Publication CIE 51 [10], amended in 1999 (Publication 51.2) [11], and published as CIE Standard S 012 in 2004 [1].

Chromaticity Limits of Daylight Simulators

As a preliminary requirement for a light source to be considered a simulator of a CIE daylight illuminant, its chromaticity coordinates must be within a specified range from the chromaticity coordinates of the illuminant. The allowable range in the CIE 1976 Uniform Chromaticity Scale diagram u_{10}', v_{10}' is a circle of radius 0.015 centered on the point representing the illuminant concerned. Figure 1 shows the allowable gamuts of chromaticity for CIE Standard Illuminant D65 and CIE illuminants D55 and D75.

CIE Method of Assessing Daylight Simulators,

Fig. 1 Allowable range of chromaticity of daylight simulator for selected CIE illuminants on the CIE 1976 Uniform Chromaticity Scale diagram u'_{10}, v'_{10}



As evaluated by the chromaticity limits both illuminants D55 and D75 are acceptable as simulators for Standard Illuminant D65, and D65 is acceptable as a simulator of both D55 and D75 (see also the category limits).

Visible Range Evaluation of Daylight Simulators

The method is based on the evaluation of the Special Metamerism Index: change in illuminants of five metameric pairs representative of practical samples in related industries. CIELAB coordinates of these sample pairs are calculated for the reference illuminant and the simulator source. The color difference between each standard and the respective comparison specimen is very near to zero for the reference illuminant; the average color difference for the five pairs under the simulator gives the quality grade for the daylight simulator.

Figure 2 illustrates the spectral radiance factors of the five standard specimens for visible range assessment based on the tables published in the CIE Standard [1].

The standard specimens are the same for every illuminant, while there are different comparison specimens for each illuminant. Figure 3 shows the

spectral radiance factors of the first standard specimen and those of the metameric comparison specimens for D50, D65 resp. D75.

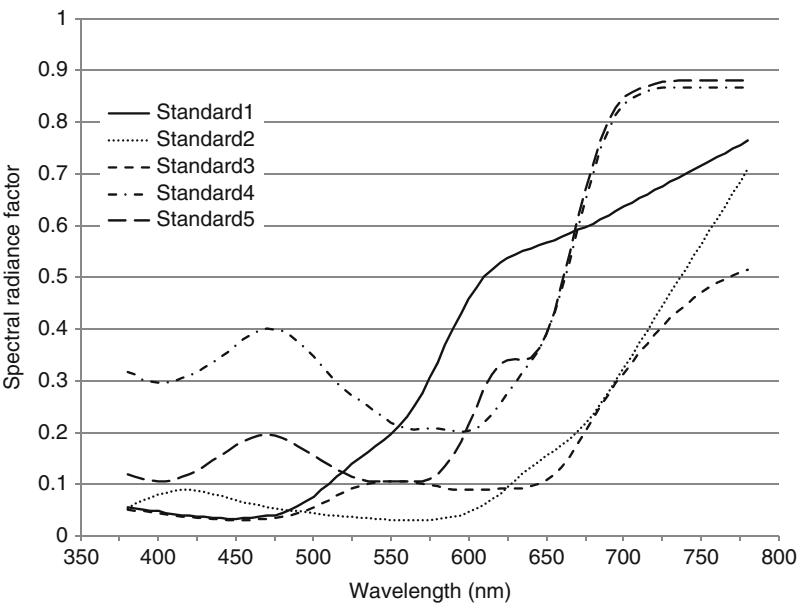
For the calculation of the visible range metamerism index M_v , the relative spectral irradiance of the simulator has to be determined in the 380–780 nm wavelength range at 5 nm intervals and over 5 nm bands and normalized so that the assessment is independent of the absolute value of irradiance. (As data are generally needed also for the determination of the UV range index, the measurements are performed, whenever viable, in the 300–780 nm range.) Tristimulus values and CIELAB coordinates are then calculated in the usual way for the five metameric pairs, and the average of the five color differences gives the visible range metamerism index M_v . The visible range quality grade is calculated according to Table 1.

The visible range quality classification of daylight simulators is thus calculated through the following steps:

- Determine the relative spectral irradiance of the light source in the 380–780 nm range.
- Calculate the tristimulus values and the CIELAB coordinates for each of the five standards and the five comparison specimens under

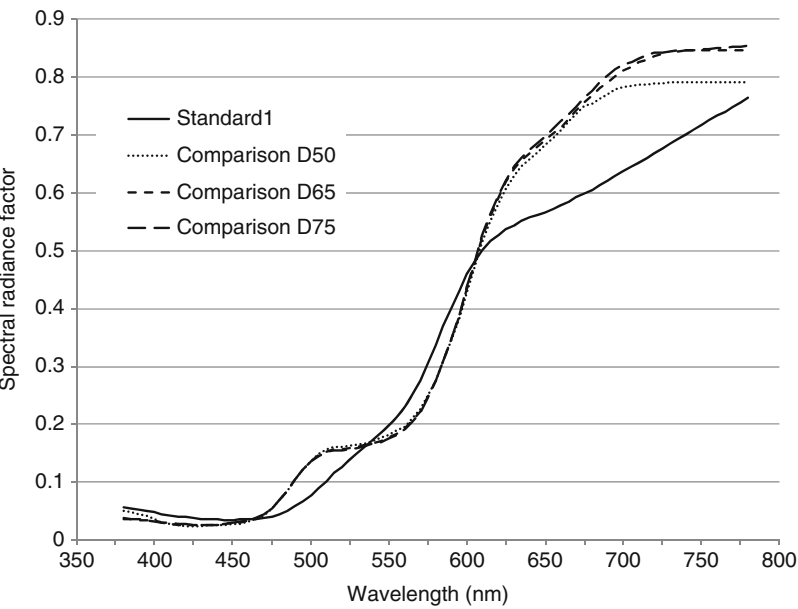
CIE Method of Assessing Daylight Simulators,

Fig. 2 Spectral radiance factors of the five standard specimens for visible range assessment



CIE Method of Assessing Daylight Simulators,

Fig. 3 Spectral radiance factors of the first standard specimen and those of the metameric comparison specimens for D50, D65 resp. D75 for visible range assessment



- the reference illuminant and under the test light source.
- Calculate the CIELAB color differences between each standard and the respective comparison specimen under reference illuminant (should be near zero) and under the test light source.
 - Calculate the average of the five color difference values.

- Determine the quality grade using Table 1.

Ultraviolet Range Evaluation of Daylight Simulators

Following the recommendation by Ganz [9], three “metameric” pairs are defined: three fluorescent standard specimens by their spectral external

radiant efficiency $Q(\lambda')$ (Fig. 4), the relative spectral distribution of radiance due to fluorescence $F(\lambda)$ (Fig. 5), and spectral reflected radiance factor $\beta_R(\lambda)$ (Fig. 6); and three nonfluorescent comparison specimens by their spectral (reflected) radiance factors (Fig. 7) (In fact, these pairs are not truly *metameric* rather *isomeric* as they are spectrally identical under the respective daylight illuminants).

The CIE document states, “ $Q(\lambda')$ is the ratio of the total radiant power emitted by the fluorescent process for an excitation wavelength λ' to the total radiant excitation power irradiating the fluorescent material” [1]. The total excitation N of the fluorescent standard specimens is computed by Eq. 1:

$$N = \sum_{300}^{460} S_n(\lambda') \cdot Q(\lambda') \cdot \Delta\lambda' \tag{1}$$

where $S_n(\lambda')$ is the normalized spectral irradiance of the simulator in the spectral region from 300 nm to 460 nm, $Q(\lambda')$ is the spectral external radiant efficiency of the fluorescent specimen over the same spectral range, as shown in Fig. 4, and $\Delta\lambda'$ is the wavelength interval of 5 nm.

$F(\lambda)$ is the ratio of the spectral distribution of radiance due to fluorescence to the sum of the tabulated values of this distribution, that is, $\sum \lambda F(\lambda) = 1.0$. $F(\lambda)$ is identical for the three fluorescent standard specimens and is independent of the SPD of the illumination.

The spectral fluorescent radiance factor $\beta_F(\lambda)$ is computed by Eq. 2:

$$\beta_F(\lambda) = \frac{N \cdot F(\lambda)}{S_n(\lambda)} \tag{2}$$

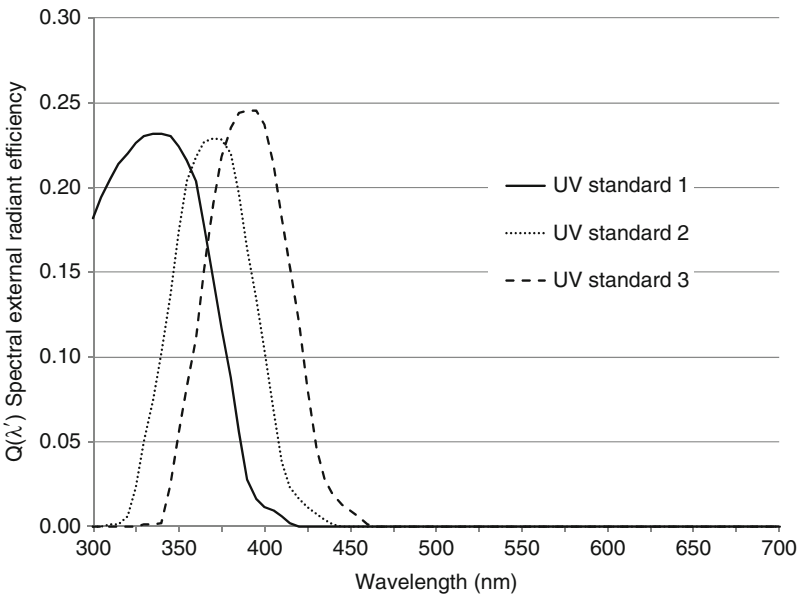
CIE Method of Assessing Daylight Simulators,
Table 1 Quality classification of daylight simulators [1]

Quality grade	Metamerism index
	Mv or Mu
A	≤ 0.25
B	$> 0.25 - 0.50$
C	$> 0.50 - 1.00$
D	$> 1.00 - 2.00$
E	> 2.00

where N is the total excitation computed by Eq. 1, $F(\lambda)$ is the relative spectral distribution of radiance due to fluorescence as shown in Fig. 5, and $S_n(\lambda)$ is the normalized spectral irradiance distribution of the simulator.

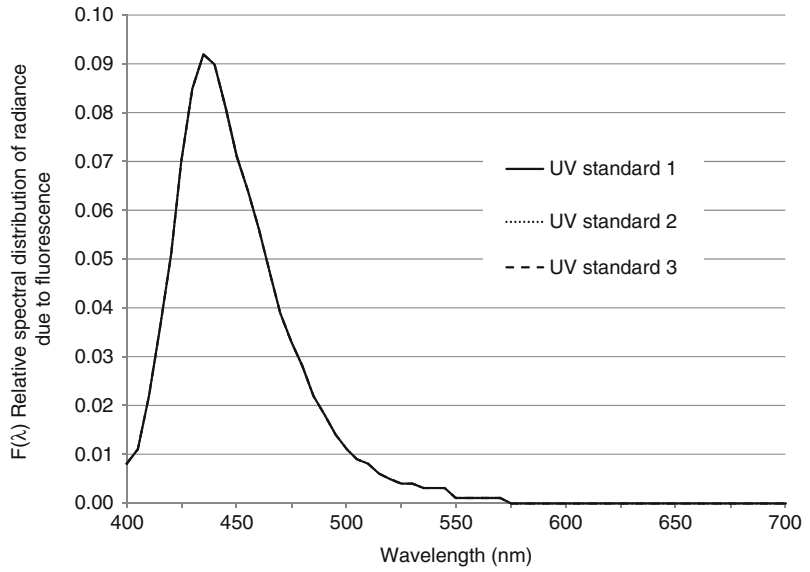
$\beta_R(\lambda)$ is the ratio of the radiance due to the reflection of the medium in the given direction to

CIE Method of Assessing Daylight Simulators,
Fig. 4 Spectral external radiant efficiency $Q(\lambda')$ for the three UV standards (based on data from [1])



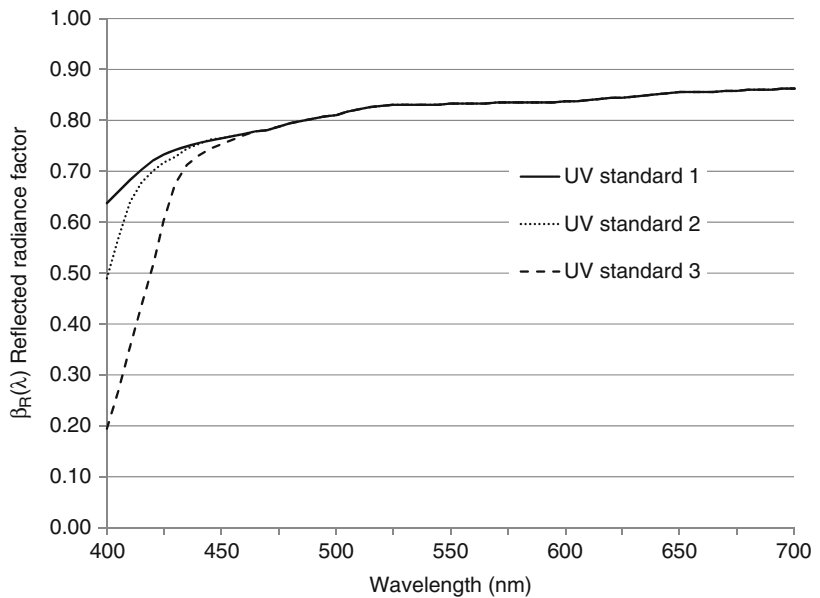
CIE Method of Assessing Daylight Simulators,

Fig. 5 Relative spectral distribution of radiance due to fluorescence $F(\lambda)$ for the three UV standards



CIE Method of Assessing Daylight Simulators,

Fig. 6 Spectral reflected radiance factor $\beta_R(\lambda)$ for the three UV standards



the radiance of a perfect reflected diffuser identically irradiated [1].

The sum of $\beta_R(\lambda)$ and $\beta_F(\lambda)$ gives the total radiance factor $\beta_T(\lambda)$:

$$\beta_T(\lambda) = \beta_R(\lambda) + \beta_F(\lambda) \quad (3)$$

as illustrated in Fig. 7.

As can be seen from the definition equations, the $\beta_F(\lambda)$ and thus the $\beta_T(\lambda)$ values depend on the

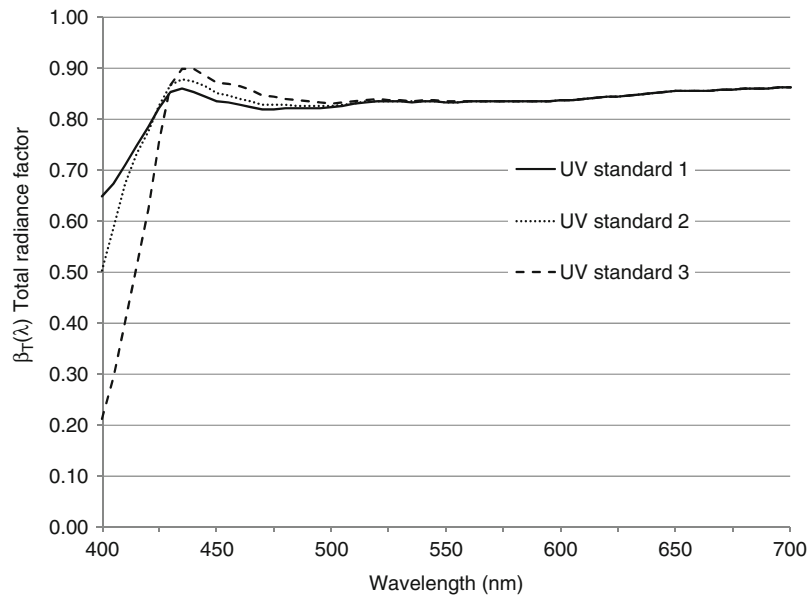
SPD of the illumination, so we would have curves for the other CIE illuminants different from those in Fig. 7.

The comparison specimens are not fluorescent; the spectral reflected radiance factors are tabulated for each CIE illuminant, as illustrated in Fig. 8 for the third specimen.

The UV standard and the UV comparison specimens are isomeric, that is, they are practically identical when illuminated by the reference

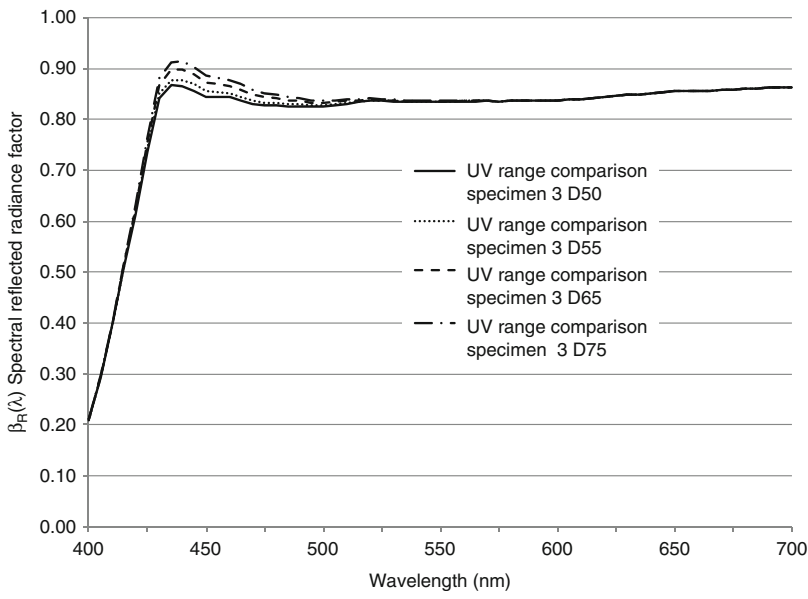
CIE Method of Assessing Daylight Simulators,

Fig. 7 Total spectral radiance factor $\beta_T(\lambda)$ for the three UV standards for illuminant D65



CIE Method of Assessing Daylight Simulators,

Fig. 8 Spectral reflected radiance factor of the first comparison specimen for the four CIE illuminants



illuminants. Thus, the UV standard 3 curve for D65 from Fig. 7 is the same as the comparison specimen 3 curve for D65 from Fig. 8. In Fig. 9, we see both curves (dotted and dashed lines) together with the standard 3 curve under a daylight simulator (solid line).

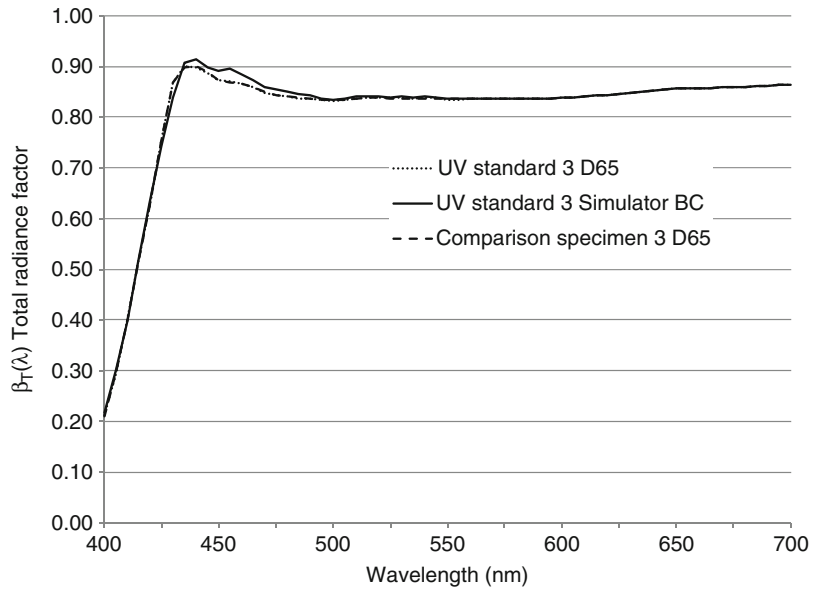
When calculating the UV range metamerism index M_u , the total radiance factor for the standard is calculated for the test source, and the spectral

(reflected) radiance factor for the comparison specimen is selected for the reference illuminant. The UV range quality classification of daylight simulators is thus calculated through the following steps:

- Determine the relative spectral irradiance of the light source $S(\lambda)$ in the 300–700 nm range.

CIE Method of Assessing Daylight Simulators,

Fig. 9 Total spectral radiance factor $\beta_T(\lambda)$ for UV standard 3 under D65 and under a quality grade BC (filtered tungsten lamp with additional UV) simulator; and for comparison specimen 3 for illuminant D65



- Take $Q(\lambda')$ for each of the three standards from the data illustrated in Fig. 4.
- Calculate N for each of the three standards from Eq. 1.
- Take the $F(\lambda)$ values for each of the three standards from the data illustrated in Fig. 5.
- Calculate $\beta_F(\lambda)$ for each of the three standards from Eq. 2.
- Take $\beta_R(\lambda)$ for each of the three standards from the data illustrated in Fig. 6.
- Calculate $\beta_T(\lambda)$ for each of the three standards from Eq. 3 as illustrated in Fig. 7.
- Take the spectral (reflected) radiance factors for the three comparison specimens (like those illustrated in Fig. 8 for specimen 3).
- Calculate the tristimulus values and the CIELAB coordinates for each of the three standards and the three comparison specimens under the reference illuminant and under the test light source.
- Calculate the CIELAB color differences between each standard and the respective comparison specimen under the reference illuminant (should be near zero) and under the test light source.
- Calculate the average of the three color difference values under the test light source.
- Determine the quality grade using Table 1.

In the example illustrated in Fig. 9, $Mu = 0$ for D65 (standard and comparison specimens are identical); $Mu = 0.79$ (quality grade C).

Practical Application of the CIE Method

In the case of color measuring spectrophotometers measuring nonfluorescent samples, the SPD of the light source has no relevance only for their visual evaluation. For fluorescent samples, both the visible and the UV range evaluation are of importance. When classifying daylight simulators, generally both quality grades are given: first the visible range metamerism index Mv , then the UV range index Mu . According to the CIE standard [1], “daylight simulators having [BC] grades have been found useful for many applications.” It is also interesting to note here that both illuminants D55 and D75 classify as grade CC simulators, while illuminant D50 is a grade DD simulator of the D65 standard illuminant.

Some national and international standards, for example, [12] also consider quality grade BC as acceptable for critical match in visual evaluations. For the classification of instruments, the ASTM Standard Practice 991–11 [13] states that the “requirement that the instrument simulation of CIE D65 shall have a rating not worse than BB

(CIELAB) as determined by the method of CIE Publication 51 has often been referenced.” This standard comes with the caveat that “the method of CIE 51 is only suitable for ultraviolet excited specimens evaluated for the CIE 1964 (10°) observer. The methods described in CIE 51 were developed for UV activated fluorescent whites and have not been proven to be applicable to visible-activated fluorescent specimens.”

There are different technologies available for realizing daylight simulators for visual assessment: filtered tungsten lamps, dichroic lamps, filtered short-arc xenon lamps, fluorescent lamps, and LED-based lamps. In color measuring spectrophotometers, filtered pulsed xenon lamps are used nearly exclusively as daylight simulator sources. A detailed description and evaluation of the different implementations was described in CIE Publication no. 192 [14] and some additional details in [15].

Many of the commercially available booths for visual assessment under D65 are quality grade BC to BE, that is, they are acceptable daylight simulators in the visible range, but only one with fluorescent lamps and one with filtered tungsten and additional UV lamps were found to be acceptable in the UV range. For D50 and D75, the results were even worse; none complied with quality grade BC criterion. Well-calibrated color measuring spectrophotometers can have excellent (quality grade AB or BA) D65 simulators, but there are no reports of instruments equipped with D50 or D75 simulators [14, 15].

References

1. CIE Standard S 012/E: Standard method of assessing the spectral quality of daylight simulators for visual appraisal or measurement of colour (2004)
2. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
3. Simonds, J.L.: Application of characteristic vector analysis to photographic and optical response data. *J. Opt. Soc. Am.* **53**, 968–971 (1963)
4. Wyszecki, G.: Development of new CIE standard sources for colorimetry. *Die Farbe*. **19**, 43–76 (1970)
5. Berger, A., Strocka, D.: Quantitative assessment of artificial light sources for the best fit to standard illuminant D65. *Appl. Opt.* **12**, 338–348 (1973)
6. Nayatani, Y., Takahama, K.: Adequateness of using 12 metameric gray object colors in appraising the color-matching properties of lamps. *J. Opt. Soc. Am.* **62**, 140–143 (1972)
7. Richter, K.: Gütebewertung der Strahldichteangleich und die Normlichtart D65. *Lichttechnik*. **24**, 370–373 (1972)
8. Berger, A., Strocka, D.: Assessment of the ultraviolet range of artificial light sources for the best fit to standard illuminant D65. *Appl. Opt.* **14**, 726–733 (1973)
9. Ganz, E.: Assessment of the ultraviolet range of artificial light sources for the best fit to standard illuminant D65. *Appl. Opt.* **16**, 806 (1977)
10. CIE: A method for assessing the quality of daylight simulators for colorimetry, publication no. 51. Central Bureau of the CIE, Vienna (1981)
11. CIE: A method for assessing the quality of daylight simulators for colorimetry, publication no. 51.2. Central Bureau of the CIE, Vienna (1999)
12. ASTM: D1729-96: Standard practice for visual appraisal of colors and color differences of diffusely-illuminated opaque materials. ASTM International, West Conshohocken (2009)
13. ASTM E991: Standard practice for color measurement of fluorescent specimens using the one-monochromator method. ASTM International, West Conshohocken (2011)
14. CIE: Practical daylight sources for colorimetry, Publication no. 192. Central Bureau of the CIE, Vienna (2010)
15. Hirschler, R., Oliveira, D.F., Lopes, L.C.: Quality of the daylight sources for industrial colour control. *Coll. Technol.* **127**, 1–13 (2011)

CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams

Andrew Stockman

Department of Visual Neuroscience, UCL
Institute of Ophthalmology, London, UK

Synonyms

CIE cone fundamentals; CIE fundamental color matching functions; Cone fundamentals, Stockman-Sharpe

Definition

Because each of the long-, middle-, and short-wavelength-sensitive (L, M, and S) cone types responds univariantly to light, human color vision and human color matches are trichromatic. Trichromatic color matches depend on the spectral sensitivities of the three cones, which are also known as the fundamental color matching functions (or CMFs): $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$. The spectral sensitivity of each cone reflects how its sensitivity changes with wavelength. Measured at the cornea, the L-, M-, and S-cone quantal spectral sensitivities peak at approximately 566, 541, and 441 nm, respectively. These fundamental CMFs are the physiological bases of other measured CMFs, all of which should be linear transformations of the fundamental CMFs.

The CIE [1] has now explicitly defined a standard set of physiologically based fundamental CMFs (or cone fundamentals) by adopting the estimates of Stockman and Sharpe [2] for 2- and 10-deg vision. These estimates were based on psychophysical measurements made in normal trichromats, red-green dichromats, blue-cone monochromats, and tritanopes all of known genotype; and from a direct analysis of the color matching data of Stiles and Burch [3].

The 10-deg cone fundamentals are defined as linear combinations of the 10-deg CMFs of Stiles and Burch [3] with some adjustments to $\bar{s}(\lambda)$ at longer wavelengths. The 2-deg cone fundamentals are similarly defined, but have also been adjusted to be appropriate for 2-deg vision.

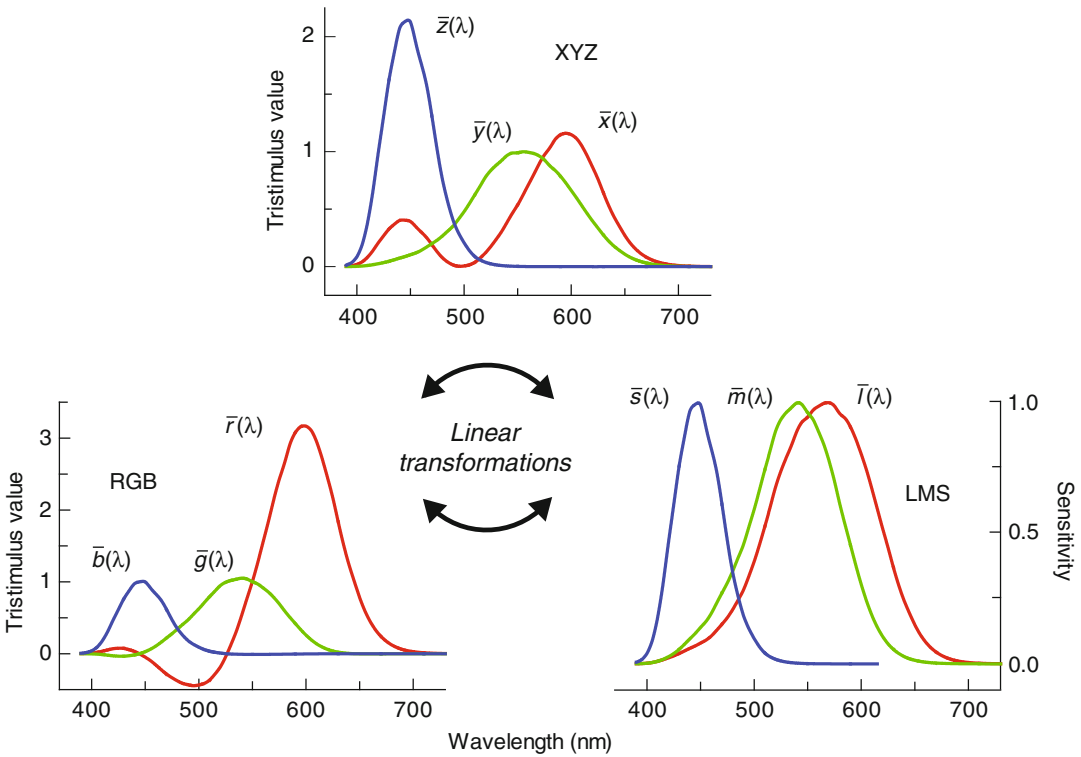
The CIE cone fundamentals are physiologically based in the sense that they reflect the spectral sensitivities of the cone photoreceptors, the initial physiological transducers of light. In principle, any set of CMFs can be linearly transformed back to the fundamental CMFs. The popular CIE 1931 CMFs, however, are substantially flawed especially at shorter wavelengths, so they cannot be used to accurately model the cone photoreceptors or indeed human color vision. One of the many advantages of using physiologically relevant functions is that they can be easily extended to represent the postreceptoral transformation of the cone signals to chromatic (L-M and S-[L + M]) and achromatic (L + M) signals.

In addition to $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$, the CIE standard also defines the photopic luminous efficiency function [$V(\lambda)$ or $\bar{y}(\lambda)$] for 2-deg and 10-deg vision as linear combinations of $\bar{l}(\lambda)$ and $\bar{m}(\lambda)$. This facilitates the further transformation of $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ to physiologically relevant versions of the more familiar CMFs: $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, for which $\bar{y}(\lambda)$ is the luminous efficiency function and $\bar{z}(\lambda)$ is a scaled version of $\bar{s}(\lambda)$.

Overview

A consequence of trichromacy is that the color of any light can be specified as the intensities of the three primary lights that match it. The bottom left-hand panel of Fig. 1 shows $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs for **RGB** (red-green-blue) primaries of 645, 526, and 444 nm. Each CMF defines the amount of that primary required to match monochromatic test lights of equal energy. CMFs, such as these, can be determined directly. CMFs can be linearly transformed to any other set of real primary lights and to imaginary primary lights, such as the **LMS** cone fundamental primaries (or “Grundempfindungen” – fundamental sensations) shown in the bottom right-hand panel of Fig. 1, which are the physiologically relevant cone spectral sensitivities, or to the still popular **XYZ** CMFs shown in the top panel. The three fundamental primaries correspond to the three imaginary primary lights that would uniquely stimulate each of the three cones and yield the $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ fundamental CMFs. All other CMF sets depend on the fundamental CMFs and should be a linear transformation of them.

A definition of the fundamental CMFs requires two things: first, an accurate set of representative $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs that can be linearly transformed to give the $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ CMFs and, second, a knowledge of the coefficients of the transformation from one to the other. Stockman and Sharpe [2] obtained the coefficients of the transformation primarily by fitting linear combinations of $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ to spectral sensitivity measurements made in red-green dichromats, blue-cone monochromats, and normals and in the case of the S-cones also by analyzing the CMFs themselves (see below).



CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 1 CMFs can be linearly transformed from one set of primaries to another.

Shown here are 10-deg CMFs for real, spectral **RGB** primaries [3] and for the CIE physiologically relevant **LMS** cone fundamental primaries and **XYZ** primaries

Choice of “Physiologically Relevant” RGB CMFs

Of critical importance in the definition of the cone fundamentals is the choice of CMFs from which they are transformed. The ones that are available vary considerably in quality. The most widely used, the CIE 1931 2-deg CMFs [4], are the least secure. Based only on the *relative* color matching data of Wright [5] and Guild [6], these CMFs were reconstructed by assuming that their linear combination must equal the 1924 CIE $V(\lambda)$ function [4, 7]. Not only is this assumption unnecessary, since CMFs can be measured directly, but the CIE $V(\lambda)$ curve used in the reconstruction is far too insensitive at short wavelengths. Thus, the CIE 1931 CMFs are a poor choice for defining the cone fundamentals.

By contrast, the Stiles and Burch 2-deg [8] and 10-deg [3] CMFs are directly measured functions. Although referred to by Stiles as “pilot” data, the 2-deg CMFs are the most extensive set of directly

measured data for 2-deg vision available, being averaged from matches made by ten observers. They are used as an intermediate step in the derivation of the cone fundamentals (see below).

The most secure and comprehensive set of directly measured color matching data are the large-field, centrally viewed 10-deg CMFs of Stiles and Burch [3]. They were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm). Consequently, the 10-deg CMFs of Stiles and Burch have been chosen as the basis for defining the “physiologically relevant” cone fundamentals. The downside of using 10-deg CMFs to model 2-deg spectral sensitivity data is that the spectral sensitivities must be corrected for the differences in preretinal filtering and in photopigment optical density between a 2-deg and 10-deg viewing field. However, such adjustments are straightforward once the spectral sensitivities are known (for details and formulae, see [9]).

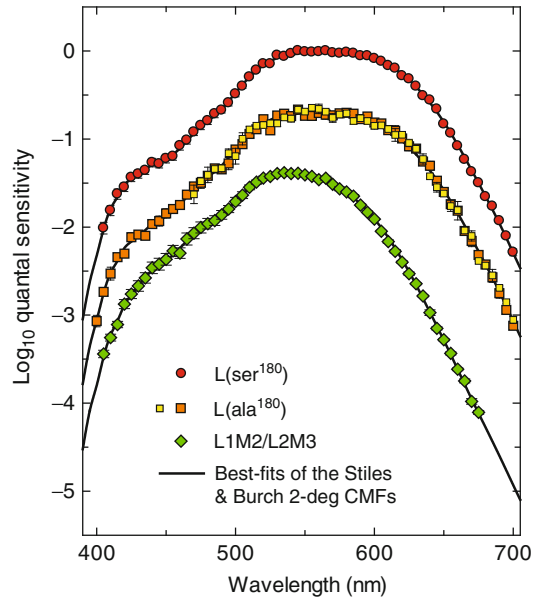
Note that the Stiles and Burch [3] 10-deg CMFs are preferable to the large-field 10-deg CIE 1964 CMFs, which, although based mainly on the 10-deg CMFs of Stiles and Burch [3], were compromised by the inclusion of the Speranskaya [10] 10-deg data and by several adjustments carried out by the CIE (see [2]).

Spectral Sensitivity Measurements

Figures 2 and 3 show the spectral sensitivity measurements from which the coefficients of the transformation to the cone fundamentals were obtained. They were measured using a 2-deg target field. Figure 2 shows the mean L- and M-cone measurements made in red-green dichromats of known genotype: deuteranopes, who lack M-cone function, either with serine (red circles) or alanine (orange and yellow squares) at position 180 of their L-cone photopigment opsin gene (these are the two commonly occurring genetic polymorphisms in the normal population that cause a slight shift in peak wavelength of the L-cone pigment), and protanopes (green diamonds), who lack L-cone. A short-wavelength chromatic adapting light eliminated any S-cone contribution to the measurements. For further details, see [11].

The upper panel of Fig. 3, below, shows the mean S-cone spectral sensitivity measurements (blue circles) made in three blue-cone monochromats, who lack L- and M-cones, and at wavelengths shorter than 540 nm in five normal subjects by Stockman, Sharpe, and Fach [12]. In normals, an intense yellow background field selectively adapted the M- and L-cones, so revealing the S-cone response at wavelengths up to 540 nm.

These spectral sensitivity measurements were then used to find the linear combinations of $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ that best fit each of the three cone spectral sensitivities, $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$, allowing adjustments in the densities of pre-receptoral filtering and photopigment optical density in order to account for differences in the mean densities between different populations and to account for differences in the retinal area (see [9]).



CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 2

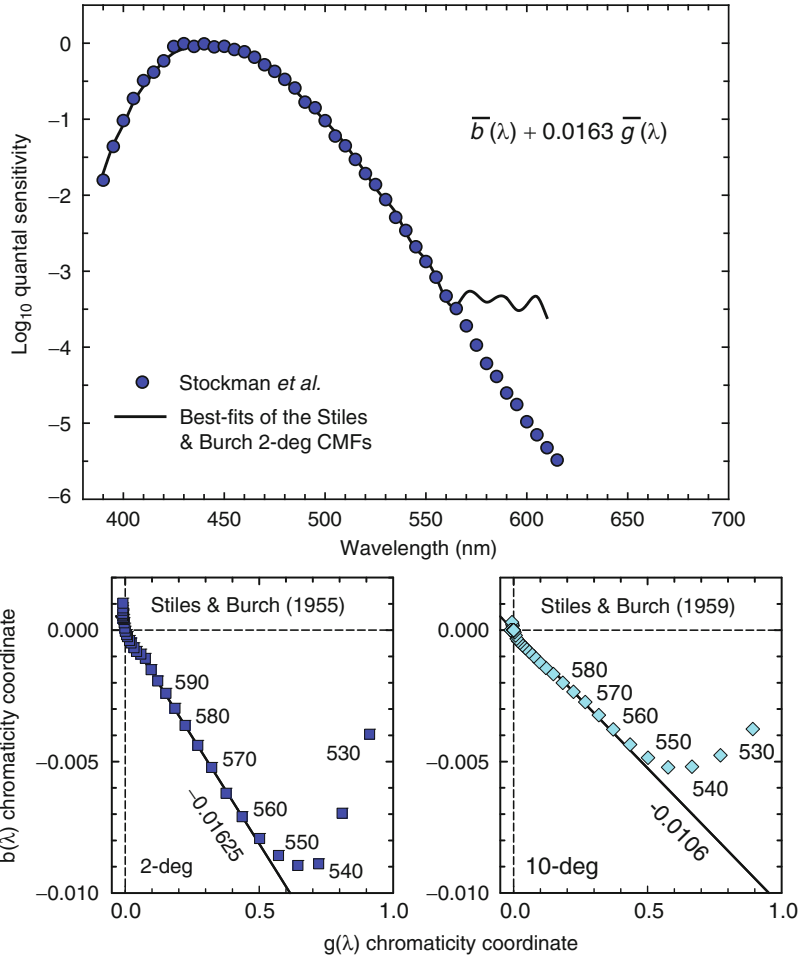
Mean cone spectral sensitivity data and fits of the CMFs. L-cone data from deuteranopes with either L(ser¹⁸⁰) (red circles, $n = 17$) or 5 L(ala¹⁸⁰) (yellow squares, $n = 2$; orange squares, $n = 3$) and M-cone data from protanopes (green diamonds, $n = 9$) measured by Sharpe et al. [11] and the linear combinations of the Stiles and Burch 2-deg CMFs [8] (continuous lines) that best fit them. The dichromat data have been adjusted in macular and lens density to best fit the CMFs. One group of L(ala¹⁸⁰) subjects did not make short-wavelength measurements. Error bars are ± 1 standard error of the mean. For best-fitting values, see Stockman and Sharpe [2]

The significance of the best-fitting linear combinations can be stated formally: When an observer matches the test and mixture fields in a color matching experiment, the two fields cause identical absorptions in each of his or her three cone types. The match, in other words, is a match at the level of the cones. The matched test and mixture fields appear identical to S-cones, to M-cones, and to L-cones. For matched fields, the following relationships apply:

$$\begin{aligned} \bar{l}_R \bar{r}(\lambda) + \bar{l}_G \bar{g}(\lambda) + \bar{l}_B \bar{b}(\lambda) &= \bar{l}(\lambda), \\ \bar{m}_R \bar{r}(\lambda) + \bar{m}_G \bar{g}(\lambda) + \bar{m}_B \bar{b}(\lambda) &= \bar{m}(\lambda), \\ \bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) &= \bar{s}(\lambda), \end{aligned} \quad (1)$$

CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams,

Fig. 3 *Top:* Mean S-cone spectral sensitivity measurements of Stockman, Sharpe, and Fach [12] and linear combination of the Stiles and Burch 2-deg CMFs that best fits them (≤ 565 nm), after applying lens and macular pigment density adjustments (blue circles). *Bottom left:* Stiles and Burch green and blue 2-deg chromaticity coordinates (blue squares). The best-fitting straight line from 555 nm to long wavelengths has a slope of -0.01625 . *Bottom right:* Stiles and Burch green and blue 10-deg chromaticity coordinates (light blue diamonds). The best-fitting straight line from 555 nm to long wavelengths has a slope of -0.0106



where $\bar{l}_R$, $\bar{l}_G$, and $\bar{l}_B$ are, respectively, the L-cone sensitivities to the R, G, and B primary lights and similarly $\bar{m}_R$, $\bar{m}_G$, and $\bar{m}_B$ and $\bar{s}_R$, $\bar{s}_G$, and $\bar{s}_B$ are the analogous L-, M-, and S-cone sensitivities. Since the S-cones are insensitive in the red part of the spectrum, it can be assumed that $\bar{s}_R$ is effectively zero for the long-wavelength R primary. There are therefore eight unknowns required for the linear transformation:

$$\begin{pmatrix} \bar{l}_R & \bar{l}_G & \bar{l}_B \\ \bar{m}_R & \bar{m}_G & \bar{m}_B \\ 0 & \bar{s}_G & \bar{s}_B \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix}. \quad (2)$$

Since we are concerned about only the relative shapes of $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$, the eight unknowns collapse to just five:

$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} k_l \bar{l}(\lambda) \\ k_m \bar{m}(\lambda) \\ k_s \bar{s}(\lambda) \end{pmatrix}, \quad (3)$$

where the absolute values of k_l (or $1/\bar{l}_B$), k_m (or $1/\bar{m}_B$), and k_s (or $1/\bar{s}_B$) remain unknown, but are typically chosen to scale three functions in some way, for example, so that $k_l \bar{l}(\lambda)$, $k_m \bar{m}(\lambda)$, and $k_s \bar{s}(\lambda)$ peak at unity.

L- and M-cone Fundamentals

The four M- and L-cone unknowns in Eq. 3, $\bar{l}_R/\bar{l}_B$, $\bar{l}_G/\bar{l}_B$, $\bar{m}_R/\bar{m}_B$, and $\bar{m}_G/\bar{m}_B$, can be estimated by fitting CMFs to the cone spectral sensitivity data shown in Fig. 2. However, since the cone spectral sensitivity data are defined for 2-deg viewing conditions and the CMFs for 10-deg, we employed an intermediate step of fitting the 2-deg data to the Stiles and Burch [8] 2-deg CMFs. Figure 2 shows the linear combinations of the Stiles and Burch 2-deg CMFs that best fit the mean L(ser^{180}) deuteranope data (red circles), L(ala^{180}) deuteranope data (yellow and orange squares), and L1M2/L2M3 protanope data (green diamonds) of Sharpe et al. [11]. An overall population mean for the L-cone spectral sensitivity function was derived by averaging the L(ser^{180}) and L(ala^{180}) fits after weighting them in ratio of 62 L(ser^{180}) to 38 L(ala^{180}), which is the ratio believed to correspond to normal population incidences (see Table 1 of Reference 2).

Having defined the mean L- and M-cone fundamentals in terms of the 2-deg Stiles and Burch CMFs, they were next defined in terms of linear combinations of the Stiles and Burch [3] 10-deg CMFs corrected to 2-deg. These were derived by a curve-fitting procedure in which the linear combinations of the Stiles and Burch 10-deg CMFs found that, after adjustment to 2-deg macular, lens and photopigment densities best fit the Stiles and Burch-based 2-deg L- and M-cone fundamentals. The coefficients are given in Eq. 4.

In one final refinement, the relative weights of the blue CMF were fine-tuned for consistency with tritanopic color matching data [13], from which the S-cones are excluded (for further details, see [2]). This final adjustment is important because of the inevitable uncertainties that arise at short wavelengths owing to individual differences in preretinal filtering.

S-Cone Fundamental

The coefficients for the transformation to the S-cone fundamental require knowledge of just one unknown, $\bar{s}_G/\bar{s}_B$, which can similarly be estimated by fitting CMFs to the cone spectral

sensitivity data. The upper panel of Fig. 3 shows the mean central S-cone spectral sensitivities (blue circles) measured by Stockman, Sharpe, and Fach [12] averaged from normal and blue-cone monochromat data below 540 nm and from blue-cone monochromat data alone from 540 to 615 nm. Superimposed on the threshold data is the linear combination of the Stiles and Burch 2-deg $\bar{b}(\lambda)$ and $\bar{g}(\lambda)$ CMFs that best fits the data below 565 nm with best-fitting adjustments to the lens and macular pigment densities.

The unknown value, $\bar{s}_G/\bar{s}_B$, can also be derived directly from the color matching data [14]. This derivation depends on the longer-wavelength part of the visible spectrum being tritanopic for lights of the radiances typically used in color matching experiments. Thus, target wavelengths longer than about 560 nm, as well as the red primary, are invisible to the S-cones. In contrast, the green and blue primaries are both visible to the S-cones. Targets longer than 560 nm can be matched for the L- and M-cones by a mixture of the red and green primaries, but a small color difference typically remains, because the S-cones detect the field containing the green primary. To complete the match for the S-cones, a small amount of blue primary must be added to the field opposite the green primary. The sole purpose of the blue primary is to balance the effect of the green primary on the S-cones. Thus, the ratio of green to blue primary should be negative and fixed at $\bar{s}_G/\bar{s}_B$, the ratio of the S-cone spectral sensitivity to the two primaries.

The lower left panel of Fig. 3 shows the Stiles and Burch [8] green, $g(\lambda)$, and blue, $b(\lambda)$, 2-deg chromaticity coordinates (blue squares). As expected, the function above ~555 nm is a straight line. It has a slope of -0.01625 , which implies $\bar{s}_G/\bar{s}_B = 0.01625$, and is the same as the value obtained from the direct spectral sensitivity measurements, 0.0163 (upper panel). The lower right panel of Fig. 3 shows the Stiles and Burch [3] green, $g(\lambda)$, and blue, $b(\lambda)$, 10-deg chromaticity coordinates and the line that best fits the data above 555 nm, which has a slope of -0.0106 (light blue diamonds). Thus, the color matching data suggest that $\bar{b}(\lambda) + 0.0106 \bar{g}(\lambda)$ is the S-cone fundamental in the Stiles and Burch [3] 10-deg space. The differences between the 2-deg

(left panel) and 10-deg (right panel) coefficients are consistent with changes in preretinal filtering and in photopigment optical density with eccentricity.

LMS Transformation Matrix

The transformation matrix from the Stiles and Burch [3] 10-deg $\bar{r}_{10}(\lambda)$, $\bar{g}_{10}(\lambda)$, and $\bar{b}_{10}(\lambda)$ CMFs to the three cone fundamentals, $\bar{l}_{10}(\lambda)$, $\bar{m}_{10}(\lambda)$, and $\bar{s}_{10}(\lambda)$ CMFs, is given by Eq. 4:

$$\begin{aligned}\bar{l}_{10}(\lambda) &= 2.846201 \bar{r}_{10}(\lambda) + 11.092490 \bar{m}_{10}(\lambda) \\ &\quad + \bar{b}_{10}(\lambda), \\ \bar{m}_{10}(\lambda) &= 0.168926 \bar{r}_{10}(\lambda) + 8.265895 \bar{g}_{10}(\lambda) \\ &\quad + \bar{b}_{10}(\lambda), \\ \bar{s}_{10}(\lambda) &= 0.010600 \bar{g}_{10}(\lambda) + \bar{b}_{10}(\lambda).\end{aligned}\quad (4)$$

The S-cone fundamental at wavelengths longer than 520 nm does not depend upon this transformation, but is based instead on the blue-cone monochromat spectral sensitivity measurements.

The 2-deg cone fundamentals are the 10-deg functions adjusted in photopigment optical density and macular pigment density according to the expected differences in those densities for 2-deg and 10-deg viewing conditions. Because the CMFs are conventionally given in energy units, this transformation yields cone fundamentals in energy units. To convert cone fundamentals in energy units to quantal units, which are more practical for vision science, multiply by λ^{-1} . The values of k_l , k_m , and k_s in Eq. 3 depend on the desired normalization and on the units (energy or quanta). For further details, see [2]. Tabulated functions can be downloaded from <http://www.cvrl.org>.

XYZ Transformation Matrix

By making a few simple assumptions, the cone fundamental CMFs, $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$, can be linearly transformed to the more familiar

colorimetric variants: $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, a form still in common use.

First, the $\bar{y}(\lambda)$ CMF is assumed to be the 2-deg or 10-deg version of the luminous efficiency functions proposed by Sharpe et al. [15], which are linear combinations of $\bar{l}(\lambda)$ and $\bar{m}(\lambda)$ (defined in Eqs. 5 and 6, below). Second, the $\bar{z}(\lambda)$ CMF is assumed to be the $\bar{s}(\lambda)$ cone fundamental scaled to have an equal integral to the $\bar{y}(\lambda)$ CMF for an equal energy white. Lastly, the definition of the $\bar{x}(\lambda)$ CMF, which owes much to the efforts of Jan Henrik Wold for the TC 1–36 committee, is based on a series of requirements: (i) like the other CMFs, the values of $\bar{x}(\lambda)$ are all positive; (ii) the integral of $\bar{x}(\lambda)$ for an equal energy white is identical to the integrals for $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$; and (iii) the coefficients of the transformation that yields $\bar{x}(\lambda)$ are optimized to minimize the Euclidian differences between the resulting $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ chromaticity coordinates and the CIE 1931 $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ chromaticity coordinates.

The 2-deg transformation is given by Eq. 5:

$$\begin{aligned}\bar{x}(\lambda) &= 1.94735469 \bar{l}(\lambda) - 1.41445123 \bar{m}(\lambda) \\ &\quad + 0.36476327 \bar{s}(\lambda), \\ \bar{y}(\lambda) &= 0.68990272 \bar{l}(\lambda) + 0.34832189 \bar{m}(\lambda), \\ \bar{z}(\lambda) &= 1.93485343 \bar{s}(\lambda),\end{aligned}\quad (5)$$

where $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ are the CIE 2-deg cone fundamentals of Stockman and Sharpe [2].

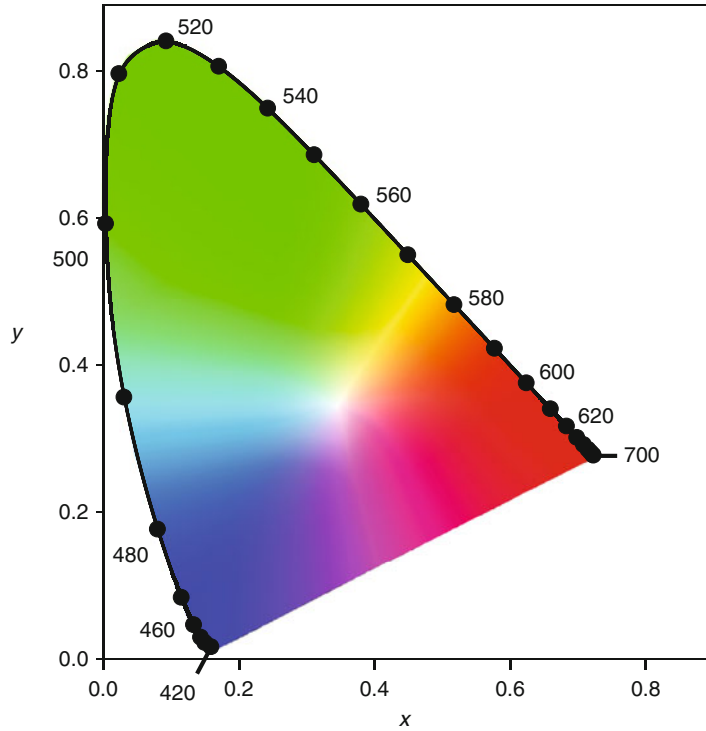
The 10-deg transformation is given by Eq. 6:

$$\begin{aligned}\bar{x}_{10}(\lambda) &= 1.93986443 \bar{l}_{10}(\lambda) - 1.34664359 \bar{m}_{10}(\lambda) \\ &\quad + 0.43044935 \bar{s}_{10}(\lambda), \\ \bar{y}_{10}(\lambda) &= 0.69283932 \bar{l}_{10}(\lambda) + 0.34967567 \bar{m}_{10}(\lambda), \\ \bar{z}_{10}(\lambda) &= 2.14687945 \bar{s}_{10}(\lambda),\end{aligned}\quad (6)$$

where $\bar{l}_{10}(\lambda)$, $\bar{m}_{10}(\lambda)$, and $\bar{s}_{10}(\lambda)$ are the CIE 10-deg cone fundamentals of Stockman and Sharpe [2]. Tabulated functions can be downloaded from <http://www.cvrl.org>.

When plotted as 2-deg chromaticity coordinates, $x(\lambda)$ and $y(\lambda)$, where:

CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 4 CIE physiologically relevant 2-deg x, y chromaticity space showing the spectrum locus (continuous line) and spectral wavelengths at every 10 nm (filled circles). An approximate representation of the color of each coordinate is shown



$$x(\lambda) = \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \quad (7)$$

- Protanopia
- Spectral luminous Efficiency
- Trichromacy

and

$$y(\lambda) = \frac{\bar{y}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)},$$

the spectrum locus and chromaticity diagram, shown in Fig. 4, have the familiar appearance of the 1931 CIE x, y chromaticity diagram.

Cross-References

- CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments
- CIE Chromaticity Coordinates (xyY)
- CIE Chromaticity Diagrams, CIE purity, CIE Dominant Wavelength
- Cone Fundamentals
- Color Categorization: Patterns and Mechanisms of Evolution
- Deuteranopia

References

1. CIE.: Fundamental chromaticity diagram with physiological axes – Part 1. Technical report pp. 170–171. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
2. Stockman, A., Sharpe, L.T.: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Res.* **40**, 1711–1737 (2000)
3. Stiles, W.S., Burch, J.M.: NPL colour-matching investigation: final report (1958). *Opt. Acta.* **6**, 1–26 (1959)
4. CIE.: Commission Internationale de l'Éclairage Proceedings, 1931. Cambridge University Press, Cambridge (1932)
5. Wright, W.D.: A re-determination of the trichromatic coefficients of the spectral colours. *Trans. Opt. Soc.* **30**, 141–164 (1928–1929); published online Epub-29
6. Guild, J.: The colorimetric properties of the spectrum. *Philos. Trans. R. Soc. Lond. A.* **230**, 149–187 (1931)
7. CIE.: Commission Internationale de l'Éclairage Proceedings, 1924. Cambridge University Press, Cambridge (1926)

8. Stiles, W.S.: Interim report to the Commission Internationale de l'Éclairage Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching (1955) with an appendix by W. S. Stiles & J. M. Burch. *Opt. Acta.* **2**, 168–181 (1955)
9. Stockman, A., Sharpe, L.T.: Cone spectral sensitivities and color matching. In: Gegenfurtner, K., Sharpe, K.T. (eds.) *Color Vision: From Genes to Perception*, pp. 53–87. Cambridge University Press, Cambridge (1999)
10. Speranskaya, N.I.: Determination of spectrum color co-ordinates for twenty-seven normal observers. *Opt. Spectrosc.* **7**, 424–428 (1959)
11. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., Nathans, J.: Red, green, and red-green hybrid photopigments in the human retina: correlations between deduced protein sequences and psychophysically-measured spectral sensitivities. *J. Neurosci.* **18**, 10053–10069 (1998)
12. Stockman, A., Sharpe, L.T., Fach, C.C.: The spectral sensitivity of the human short-wavelength cones. *Vision Res.* **39**, 2901–2927 (1999); published online EpubAug
13. Stockman, A., Sharpe, L.T.: Tritanopic color matches and the middle- and long-wavelength-sensitive cone spectral sensitivities. *Vision Res.* **40**, 1739–1750 (2000)
14. Bongard, M.M., Smirnov: M.S.: Determination of the eye spectral sensitivity curves from spectral mixture curves. *Doklady Akademii nauk S.S.S.R.* **102**, 1111–1114 (1954)
15. Sharpe, L.T., Stockman, A., Jagla, W., Jägle, H.: A luminous efficiency function, $V^*(\lambda)$, for daylight adaptation: a correction. *Color Res. Appl.* **36**, 42–46 (2011)

CIE Special Metamerism Index: Change in Observer

Abhijit Sarkar
Mixed Reality Systems Engineering, Microsoft
Corporation, Redmond, WA, USA

Synonyms

[CIE Standard Deviate Observer](#)

Definition

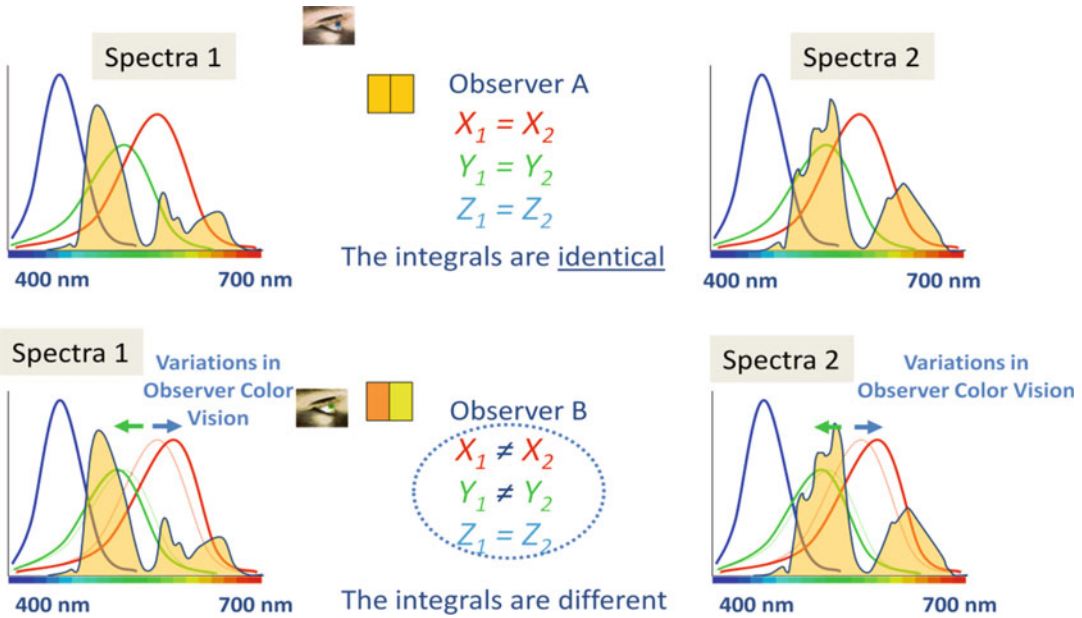
Special Metamerism Index: Change in observer refers to a method for evaluating the average of

and the range of color mismatches for metameric color pairs when test observers with normal color vision are substituted for a reference observer (i.e., a CIE standard colorimetric observer). This method was proposed in 1989 by the CIE Technical Committee 1-07 [1]. The method, principally based on the works of Nayatani et al. [2] and Takahama et al. [3], was intended to evaluate both the average values and the range of color mismatches.

Overview

Two color stimuli with different spectral characteristics within the visible spectral range can have identical tristimulus values for a given illuminant and the reference observer (standard colorimetric observer). These stimuli are said to be metameric. Such metameric color match breaks down on changing the illuminant (illuminant metamerism) or the observer (observer metamerism). This is depicted in Fig. 1. While for observer A (reference observer), spectra 1 and spectra 2 (shaded curves) result in identical tristimulus values, they are no longer identical in the case of observer B (test observer), since this observer's color-matching functions differ from those of observer A. If the spectral characteristics of the primary colorants of two-color reproduction devices are not the same, any color match made on these devices is metameric in nature and thus may not hold when one observer is replaced by another.

Given the fact that the concept of metamerism is fundamentally important in the science of colorimetry, it was recognized long ago that a suitable index to quantify this effect would be of high relevance. It is relatively straightforward to formulate specific indices for illuminant metamerism, based on the color differences of a metameric pair under various sources of illumination. However, it is quite challenging to define a general index for observer metamerism. This is due to the immense variability that exists in the spectral sensitivities of the cone photoreceptors of individual observers having normal color vision. This variability is manifested in the measured color-matching functions and thus in color



CIE Special Metamerism Index: Change in Observer, Fig. 1 A graphical depiction of the phenomenon of observer metamerism. For observer A (the reference observer), the tristimulus values resulting from spectrally integrating the color-matching functions of observer A and

the two spectra are identical. However, that is no longer the case for observer B, whose color-matching functions are different from those of observer A. Thus, spectra 1 and 2 are metameric. (Image courtesy: Dr. Laurent Blondé)

mismatches in a metameric pair when evaluated by individual observers.

Observer metamerism can be evaluated using experimental procedure involving actual observers with normal color vision or through mathematical modeling, using a set of color-matching functions that is representative of individual variations among color-normal observers. In one of the first attempts to model the uncertainties involved in the color-matching data, Nimeroff et al. [4] proposed a statistical model they termed as complete standard colorimetric observer system. The model included the mean of the color-matching functions of various observers as well as variance and covariance of these functions derived from the intra- and interobserver variability. On the other hand, Wyszecki and Stiles [5] attempted to define an index for observer metamerism by using the color-matching functions of 20 individual observers from the large-field color-matching

experiment of Stiles and Burch [6]. They computed color differences for a given metameric pair as perceived by each of the 20 observers and used the mean color difference as the degree of observer metamerism. However, this method was not proposed with industrial applications in mind.

In 1989, a method was formulated by the CIE for evaluating observer metamerism. The method was detailed in a technical report titled “CIE Special Metamerism Index: Change in Observer,” prepared by the CIE Technical Committee (TC) 1-07. The committee came under CIE Division 1 (Vision and Color) and was chaired by Professor N. Ohta. This method was analogous to the one for evaluating illuminant metamerism, proposed by the CIE in 1986: “Special Metamerism Index: Change in Illuminant” [7]. This method, described hereafter as CIE observer metamerism index, assessed the degree of color mismatch for metameric color pairs (object colors

or illuminant colors), resulting from substitution of the reference observer by a test observer. The reference observer is either the CIE 1931 standard colorimetric observer or the CIE 1964 supplementary standard colorimetric observer. The test observers are assumed to be a number of actual observers with normal color vision and are represented by four deviation functions characterizing the variations of color-matching functions of color-normal observers.

Standard Deviate Observer

Computation of the CIE observer metamerism index requires the use of color-matching functions for a standard deviate observer. Thus, the definition of the standard deviate observer and an overview of various propositions for its formulation would be relevant here.

The concept of standard deviate observer is a mathematical construct that was first proposed by Allen in 1970 [8]. This observer has color-matching functions differing from the reference observer by amounts equal to standard deviations among a defined set of color-normal observers. Allen and subsequently other researchers used the 20 individual Stiles and Burch observers [6] for deriving the standard deviate observer.

In the method proposed by Allen [8], starting from the standard deviations $\Delta\bar{x}(\lambda)$, $\Delta\bar{y}(\lambda)$, and $\Delta\bar{z}(\lambda)$ of color-matching functions of the 20 Stiles and Burch observers, differential tristimulus values ΔX , ΔY , and ΔZ were computed for any metameric pair. These values represented the root mean square differences in tristimulus values for a metameric pair as perceived by the observers. The tristimulus and delta tristimulus values could then be used in a color-difference formula to calculate ΔE , which was the desired observer metamerism index. Allen's method allowed for negative differences in tristimulus values.

In a different statistical approach, Nayatani et al. [9] performed a singular value decomposition analysis on the 20-observer data and derived three deviation functions whose linear combinations were used to reconstitute the color-matching

functions of the 20 Stiles and Burch observers. Further, the authors showed that the first deviation function was similar to Allen's standard deviate observer. Adding the first deviation function to the reference observer (e.g., CIE 1964 supplementary standard colorimetric observer) yielded the test observer. Observer metamerism indices derived by using the first deviation function showed high correlation to the average metamerism index of the 20 Stiles and Burch observers. These results were obtained by using two sets of metameric spectral reflectance values of 12 and 68 metamers.

A subsequent study by Takahama et al. [3] expanded the method by using the first deviation to evaluate the index of observer metamerism. All four deviation functions were used to construct the confidence ellipsoids of tristimulus values defining the range of mismatches expected for a given pair of metamers, viewed by actual observers with normal color vision but different from the reference.

In an independent study, Ohta [10] performed a nonlinear optimization of the 20-observer data to formulate a standard deviate observer model. The model was close to the one obtained by Nayatani [9] and was assessed to well represent the original 20 observers.

Procedure for Computing CIE Observer Metamerism Index

Tristimulus values for a pair of object colors metameric for a reference observer with color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ and an illuminant with spectral power distribution $S(\lambda)$ are given by Eq. 1.

$$\begin{bmatrix} X_{ref,i} \\ Y_{ref,i} \\ Z_{ref,i} \end{bmatrix} = \sum_{\lambda} \rho_i(\lambda) S(\lambda) \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} \Delta\lambda \quad (1)$$

where $\rho_i(\lambda)$ is the spectral reflectance of the i -th object color ($i = 1, 2$) of the metameric pair.

Since the object colors are metameric, this can be written as:

$$\begin{aligned} X_{ref,1} &= X_{ref,2} \cong X_{ref} \\ Y_{ref,1} &= Y_{ref,2} \cong Y_{ref} \\ Z_{ref,1} &= Z_{ref,2} \cong Z_{ref} \end{aligned} \quad (2)$$

Further, $S(\lambda)$ is normalized so that it has a luminance of 100, as shown in Eq. 3.

$$\sum_{\lambda} S(\lambda) \bar{y}(\lambda) \Delta\lambda = 100 \quad (3)$$

Four sets of deviation functions $\bar{x}_j(\lambda)$, $\bar{y}_j(\lambda)$, and $\bar{z}_j(\lambda)$, where j denotes the set number, are used in this method and have been tabulated in the report of CIE TC 1-07.

The first set of deviation functions denotes the differences in color-matching functions of CIE standard colorimetric observer and the standard deviate observer [denoted by $\bar{x}_{dev}(\lambda)$, $\bar{y}_{dev}(\lambda)$, and $\bar{z}_{dev}(\lambda)$]. Thus, the color-matching functions of the standard deviate observer are obtained by using Eq. 4.

$$\begin{aligned} \bar{x}_{dev}(\lambda) &= \bar{x}(\lambda) + \Delta\bar{x}_1(\lambda) \\ \bar{y}_{dev}(\lambda) &= \bar{y}(\lambda) + \Delta\bar{y}_1(\lambda) \\ \bar{z}_{dev}(\lambda) &= \bar{z}(\lambda) + \Delta\bar{z}_1(\lambda) \end{aligned} \quad (4)$$

From these color-matching functions, the tristimulus values $X_{dev,i}$, $Y_{dev,i}$, and $Z_{dev,i}$ of metameric object colors ($i = 1$ or 2) corresponding to the standard deviate observer can be obtained as shown in Eq. 5.

$$\begin{bmatrix} X_{dev,i} \\ Y_{dev,i} \\ Z_{dev,i} \end{bmatrix} = \sum_{\lambda} \rho_i(\lambda) S(\lambda) \begin{bmatrix} \bar{x}_{dev}(\lambda) \\ \bar{y}_{dev}(\lambda) \\ \bar{z}_{dev}(\lambda) \end{bmatrix} \Delta\lambda \quad (5)$$

The CIE observer metamerism index (M_{obs}) for the pair of metameric object colors is expressed by Eq. 6.

$$M_{obs} = \Delta E_{obs}^* \left[(X_{dev,1}, Y_{dev,1}, Z_{dev,1}), (X_{dev,2}, Y_{dev,2}, Z_{dev,2}) \right] \quad (6)$$

where ΔE_{obs}^* is the color difference between the metameric object colors as evaluated by the standard deviate observer, calculated in a uniform

color space. CIE TC 1-07 recommends CIE 1976 (L^* , u^* , v^*) or (L^* , a^* , b^*) as uniform color space [7].

When the tristimulus values of the samples do not exactly match, they are not strictly metameric (instead, they are parameric). In such a scenario, the tristimulus values of the first stimulus are defined as reference (X_{ref} , Y_{ref} , and Z_{ref}), as in Eq. 7.

$$\begin{aligned} X_{ref} &= X_{ref,1} (\neq X_{ref,2}) \\ Y_{ref} &= Y_{ref,1} (\neq Y_{ref,2}) \\ Z_{ref} &= Z_{ref,1} (\neq Z_{ref,2}) \end{aligned} \quad (7)$$

The tristimulus values of the second stimulus obtained by Eq. 5 are then corrected using Eq. 8.

$$\begin{aligned} X'_{dev,2} &= X_{dev,2} \left(\frac{X_{ref,1}}{X_{ref,2}} \right) \\ Y'_{dev,2} &= Y_{dev,2} \left(\frac{Y_{ref,1}}{Y_{ref,2}} \right) \\ Z'_{dev,2} &= Z_{dev,2} \left(\frac{Z_{ref,1}}{Z_{ref,2}} \right) \end{aligned} \quad (8)$$

Deriving 95% Confidence Ellipse

Resultant tristimulus values of metameric colors inevitably mismatch when evaluated by a test observer. These tristimulus values spread within a certain range in the three-dimensional color space. This range is characterized in a chromaticity diagram by a statistical confidence ellipse encompassing 95% of the spread. In this way, evaluation of observer metamerism allows an estimation of tolerances in color-difference judgments for various metameric color pairs.

All four deviate observer functions are used to estimate the confidence ellipse containing 95% of color mismatches. First, a set of tristimulus value deviations for a pair of metameric object colors is defined using Eq. 9.

$$\begin{bmatrix} \Delta^2 X_j \\ \Delta^2 Y_j \\ \Delta^2 Z_j \end{bmatrix} = \sum_{\lambda} [\rho_2(\lambda) - \rho_1(\lambda)] S(\lambda) \begin{bmatrix} \bar{x}_j(\lambda) \\ \bar{y}_j(\lambda) \\ \bar{z}_j(\lambda) \end{bmatrix} \Delta\lambda \quad (9)$$

Here, the spectral reflectance factors of the object colors are represented by $\rho_1(\lambda)$ and $\rho_2(\lambda)$, and j refers to one of the four deviation functions. As before, the spectral power distribution of the illuminant is denoted by $S(\lambda)$.

In case of metameric illuminant colors, the spectral power distributions $S_1(\lambda)$ and $S_2(\lambda)$ of the illuminant pair are normalized similar to Eq. 3. Instead of Eqs. 9 and 10 is used.

$$\begin{bmatrix} \Delta^2 X_j \\ \Delta^2 Y_j \\ \Delta^2 Z_j \end{bmatrix} = \sum_{\lambda} [S_2(\lambda) - S_1(\lambda)] \begin{bmatrix} \bar{x}_j(\lambda) \\ \bar{y}_j(\lambda) \\ \bar{z}_j(\lambda) \end{bmatrix} \Delta\lambda \quad (10)$$

Next, two matrices D and V are defined as shown in Eqs. 11 and 12.

$$D = \frac{1}{(X_{ref} + 15Y_{ref} + 3Z_{ref})^2} \begin{bmatrix} 60Y_{ref} + 12Z_{ref} & -60X_{ref} & -12X_{ref} \\ -9Y_{ref} & 9X_{ref} + 27Z_{ref} & -27Y_{ref} \end{bmatrix} \quad (11)$$

$$V = \sum_{i=1}^4 \begin{bmatrix} (\Delta^2 X_i)^2 & \Delta^2 X_i \cdot \Delta^2 Y_i & \Delta^2 Z_i \cdot \Delta^2 X_i \\ \Delta^2 X_i \cdot \Delta^2 Y_i & (\Delta^2 Y_i)^2 & \Delta^2 Y_i \cdot \Delta^2 Z_i \\ \Delta^2 Z_i \cdot \Delta^2 X_i & \Delta^2 Y_i \cdot \Delta^2 Z_i & (\Delta^2 Z_i)^2 \end{bmatrix} \quad (12)$$

The variance-covariance matrix Σ of the (u', v') chromaticity coordinates of all the color matches evaluated by test observers is given by Eq. 13.

$$\Sigma = DVD^t \quad (13)$$

where D^t is the transpose of matrix D . If the elements of inverse matrix Σ^{-1} are denoted as Σ_{mn} where m is the row and n is the column of a given element, the confidence ellipse containing 95% of the color mismatches is given by Eq. 14.

$$\begin{aligned} & \sum^{11} (\Delta u')^2 + 2 \sum^{12} (\Delta u') (\Delta v') \\ & + \sum^{22} (\Delta v')^2 \\ & = \chi^2(2, 0.05) = 5.991 \end{aligned} \quad (14)$$

Here, $\chi^2(2, 0.05)$ is the 5% of the χ^2 distribution for two degrees of freedom.

The center of the ellipse is given by (u_{ref}, v_{ref}) , which corresponds to the reference observer.

It should be noted that the two-dimensional ellipse in the chromaticity diagram does not contain information about the psychometric lightness of the stimuli. Thus, the ellipse cannot be used for comparing the degree of observer metamerism between two different metameric pairs with varying lightness.

Effect of Age on Color Mismatches

The CIE TC 1-07 report [1] also proposes a method that accounts for the effect of age on the index of observer metamerism. Mainly, aging of the lens (pigmentation) is considered as a contributing factor. A pair of object colors which are metameric for an average observer with an age N_1 and a reference illuminant will mismatch for an observer with an age N_2 . The deviation functions $\Delta\bar{x}_1(\lambda, N)$, $\Delta\bar{y}_1(\lambda, N)$, and $\Delta\bar{z}_1(\lambda, N)$ for age N ($20 \leq N \leq 60$) are computed using Eqs. 15 and 16. These values are used instead of those obtained from Eq. 4.

$$\begin{bmatrix} \Delta\bar{x}_1(\lambda, N) \\ \Delta\bar{y}_1(\lambda, N) \\ \Delta\bar{z}_1(\lambda, N) \end{bmatrix} = L(N) \cdot \begin{bmatrix} \Delta\bar{x}_1(\lambda) \\ \Delta\bar{y}_1(\lambda) \\ \Delta\bar{z}_1(\lambda) \end{bmatrix} \quad (15)$$

$$L(N) = 0.064 \cdot N - 2.31 \quad (16)$$

Evaluation of the CIE Observer Metamerism Index

Following the introduction of CIE observer metamerism index [1], several researchers evaluated the model with independent experimental data. North and Fairchild [11] conducted a Maxwell-type color-matching experiment using a cathode ray tube (CRT) display and a tungsten-halogen lamp on two halves of a 2° bipartite field. The authors estimated the color-matching functions of each observer through a mathematical model, starting from experimental data obtained at seven wavelengths. They concluded that the interobserver variability in the experimental data was much larger than what was predicted by the CIE model [1].

Alfvin and Fairchild [12] conducted another visual experiment on color matches between color prints or transparencies and a CRT display. They used an equilateral glass prism to allow the observers to view simultaneously both the soft and hard copy stimuli in a vertically symmetric bipartite field. In analyzing the data, they arrived at the same conclusion: The interobserver variability was significantly larger than the prediction

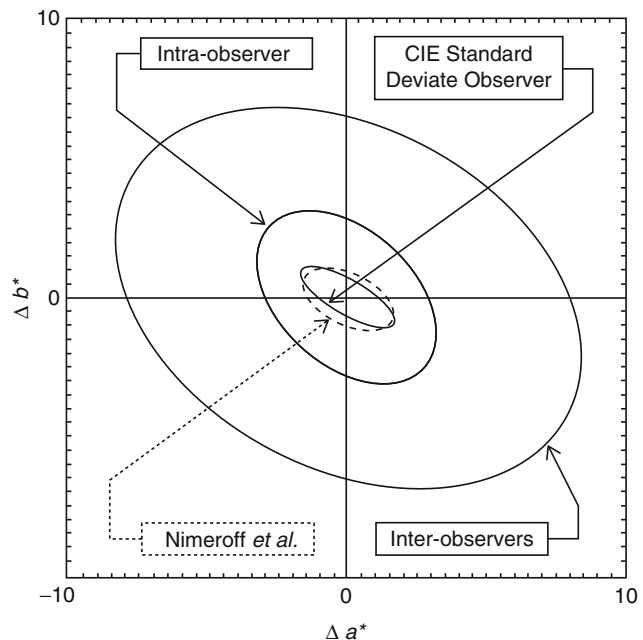
of the CIE observer metamerism index [1]. As an example, Fig. 2 shows bivariate 95% confidence ellipses containing the range of CIELAB (Δa^* , Δb^*) color mismatches for cyan-transparency sample. The plot includes bivariate ellipse calculated using the CIE method [1] as well as experimentally determined bivariate ellipses for intra- and interobserver color matches. According to these results, the interobserver variability is significantly underpredicted by the CIE observer metamerism index [1].

Oicherman et al. [13] conducted an asymmetric color-matching experiment where 11 observers were asked to match the colors displayed on a CRT and an LCD to the colors of 2 achromatic and 8 chromatic paint samples placed inside a light booth. Even in this study, the authors reported a significant underprediction of the observer variations of color-matching data by the CIE observer metamerism index [1], accounting for only 15% of interobserver variability.

Why did the CIE observer metamerism index not perform well? The suggested explanations include the exclusion of some of the Stiles and Burch observers from the analysis that led to the development of the CIE observer metamerism

CIE Special Metamerism Index: Change in Observer,

Fig. 2 Comparison of 95% confidence regions for measured and predicted ranges of color mismatch for the cyan transparency. Measured data are shown for Alfvin and Fairchild [12] and Nimeroff et al. [4], along with confidence region predicted by the CIE observer metamerism index (CIE standard deviate observer). All data are shown in CIELAB $\Delta a^* - \Delta b^*$ plane. (Reproduced from Alfvin and Fairchild [12])



index [11] and improper mathematical treatment of the original color-matching data [12]. Looking from the point of view of practical industrial applications, in particular hard copy versus soft copy color matching, some researchers [13] have questioned the purpose and usefulness of an index of observer metamerism and a standard deviate observer. They suggested that individual variability in these conditions is governed by mechanisms of chromatic discrimination and could be modeled by advanced color-difference formulas with suitably adjusted parametric coefficients.

Future Directions

In 2006, CIE's Technical Committee 1-36 published a report [14] on the choice of a set of color-matching functions and estimates of cone fundamentals for the color-normal observer. Starting from the 10° color-matching functions of 47 observers from Stiles and Burch's large-field color-matching experiment [6], the model defines 2° and 10° fundamental observers and provides a convenient framework for calculating average cone fundamentals for any field size between 1° and 10° and for an age between 20 and 80. This model provides a theoretical framework for quantifying observer metamerism resulting from age variations. Subsequently, CIE TC 170-1 published guidance [15] to mathematically compute spectral luminous efficiency functions and chromaticity diagrams for physiological observers.

Adopting a different approach to address the issue of observer metamerism in applied colorimetry, Sarkar et al. [16] introduced the concept of colorimetric observer categories. In this work, eight of such categories were derived through statistical analyses of a combined dataset of experimental and physiological color-matching functions. Physiological data were obtained from the mathematical model proposed by CIE TC 1-36 [14], while experimental data consisted of color-matching functions of 47 Stiles and Burch observers [6]. The authors developed and implemented an experimental method to classify

a color-normal human observer as belonging to one of these categories, based on his or her color vision. Subsequently, a compact proof-of-concept prototype for conducting such experiments was developed [17].

As a follow-up to that work, Asano et al. developed a synthetic vision model for individual colorimetric observers [18]. The authors extended the CIE TC 1-36 physiological observer by adding eight additional physiological parameters. Corresponding population variability was defined by standard deviations, which was then used to simulate a large dataset of color-matching functions using Monte Carlo simulation. Subsequently, Asano and Fairchild proposed a set of categorical observer functions [19]. A similar methodology as that of Sarkar et al. [17] was used, but there were a couple of key differences in the approach. While Sarkar et al.'s method relied on measured color-matching functions of Stiles-Burch observers and CIE TC 1-36 model prediction essentially based on the same measured data, Asano and Fairchild's method used synthetic vision model developed earlier by Asano et al. [18]. Further, Sarkar et al. envisioned that a unique set of observer categories could be established using a standardized apparatus based on large field-of-view (10°) color matching. Asano and Fairchild's observer models on the other hand are based entirely on theoretical analysis of the synthetic dataset, to avoid any dependence on the spectral properties of color primaries of any specific device. A set of ten categorical observers was found to be optimal for general use.

At the time of writing this essay, more and more studies are leveraging the concept of observer categories in investigating various aspects of observer metamerism, such as the one by Park et al. [20]. As various researchers continue to validate the usefulness of the concept of observer categories and observer classification in the study of observer metamerism, this can eventually lead to an alternative method for quantifying such metamerism in applied colorimetry as well as for providing a practical solution to the issue of observer metamerism in various color-critical industrial applications.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [Metamerism](#)

References

1. CIE Publication 80: Special Metamerism Index: Change in Observer. Central Bureau of the CIE, Vienna (1989)
2. Nayatani, Y., Takahama, K., Sobagaki, H.: A proposal of new standard deviate observers. *Color. Res. Appl.* **8**(1), 47–56 (1983)
3. Takahama, K., Sobagaki, H., Nayatani, Y.: Prediction of observer variation in estimating colorimetric values. *Color. Res. Appl.* **10**(2), 106–117 (1985)
4. Nimeroff, I., Rosenblatt, J.R., Dannemiller, M.C.: Variability of spectral tristimulus values. *J. Opt. Soc. Am. A*. **52**, 685–691 (1962)
5. Wyszecki, G., Stiles, W.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. Wiley, New York (1982)
6. Stiles, W., Burch, J.: NPL colour-matching investigation: final report (1958). *J. Mod. Opt.* **6**, 1–26 (1959)
7. CIE Publication 15-2: Colorimetry, 2nd edn. Central Bureau of the CIE, Vienna (1986)
8. Allen, E.: An index of metamerism for observer differences. In: *Proceedings of the 1st AIC Congress, Color 69*, Musterschmidt, Göttingen, pp. 771–784 (1970)
9. Nayatani, Y., Hashimoto, K., Takahama, K., Sobagaki, H.: Comparison of methods for assessing observer metamerism. *Color. Res. Appl.* **10**, 147–155 (1985)
10. Ohta, N.: Formulation of a standard deviate observer by a nonlinear optimization technique. *Color. Res. Appl.* **10**, 156–164 (1985)
11. North, A., Fairchild, M.: Measuring color-matching functions. Part II. New data for assessing observer metamerism. *Color. Res. Appl.* **18**, 163–170 (1993)
12. Alfvén, R., Fairchild, M.: Observer variability in metameric color matches using color reproduction media. *Color. Res. Appl.* **22**, 174–188 (1997)
13. Oicherman, B., Luo, M., Tarrant, A., Robertson, A.: The uncertainty of colour-matching data. In: *Proceedings of IS&T/SID 13th Color Imaging Conference*, pp. 326–332 (2005)
14. CIE 170-1:2006 Fundamental Chromaticity Diagram with Physiological Axes – Part 1. (2006)
15. CIE 170-2:2015 Fundamental Chromaticity Diagram with Physiological Axes – Part 2: Spectral Luminous Efficiency Functions and Chromaticity Diagrams. (2015)
16. Sarkar, A., Blondé, L., Le Callet, P., Autrusseau, F., Stauder, J., Morvan, P.: Toward reducing observer metamerism in industrial applications: colorimetric observer categories and observer classification. In: *Proceedings of the 18th Color and Imaging Conference, San Antonio* (2010)
17. Sarkar, A.: Identification and assignment of colorimetric observer categories and their applications in color and vision sciences. PhD thesis, Université de Nantes (2011)
18. Asano, Y., Fairchild, M.D., Blondé, L.: Individual colorimetric observer model. *PLoS One*. **11**(2), e0145671 (2016)
19. Asano, Y., Fairchild, M.D.: Categorical observers for metamerism. *Color. Res. Appl.* **45**, 576–585 (2020)
20. Park, Y., Murdoch, M.J., Fairchild, M.D.: Observer metamerism: why do [mis]matches of neutral appear pinkish or greenish. In: *Proceedings of the 28th Color and Imaging Conference*, pp. 7–12 (2020)

CIE Standard Deviate Observer

- [CIE Special Metamerism Index: Change in Observer](#)

CIE Standard Illuminant A

- [CIE Standard Illuminants and Sources](#)

CIE Standard Illuminant D65

- [CIE Standard Illuminants and Sources](#)

CIE Standard Illuminants and Sources

János Schanda
Veszprém, Hungary

Synonyms

[CIE standard illuminant A](#); [CIE standard illuminant D65](#); [CIE standard source A](#)

János Schanda: Deceased.

Definitions

CIE Standard Illuminants

Illuminants A and D65 defined by the CIE in terms of relative spectral power distributions [1].

Note 1: These illuminants are intended to represent:

A: Planckian radiation at a temperature of about 2,856 K

D65: The relative spectral power distribution representing a phase of daylight with a correlated color temperature of approximately 6,500 K (called also “nominal correlated color temperature of the daylight illuminant”)

Note 2: Illuminants B and C and other D illuminants, previously denoted as “standard illuminants,” should now be termed “CIE illuminants.”

CIE Standard Sources

Artificial sources specified by the CIE whose radiation approximate CIE standard illuminants.

Note: CIE sources are artificial sources that represent CIE illuminants.

Overview

As shown already in the section Definitions, one has to distinguish between illuminants and sources: The term “illuminant” refers to a defined spectral power distribution, not necessarily realizable or provided by an artificial source. Illuminants are used in colorimetry to compute the tristimulus values of reflected or transmitted object colors under specified conditions of illumination. The term “source” refers to a physical emitter of light, such as a lamp or the sky.

The CIE colorimetric system was established in 1931 [2], and CIE defined in those days three illuminants: illuminants A, B, and C. Illuminant A was chosen to resemble the spectral power distribution (SPD) of an average incandescent lamp, the SPD of illuminant B was near to that of average direct sunlight, and illuminant C represented average daylight. During the years, it turned out that illuminant B was very seldom used and was soon dropped. Illuminant C was defined only in the visible

spectrum, and with the introduction of optical brighteners, colorimetry needed illuminants with defined ultraviolet radiation content. In 1964 the CIE recommended a new set of daylight illuminants [3], where the SPD was also defined in the ultraviolet (UV) part of the spectrum, and decided that one of these, with a correlated color temperature near to 6,500 K, should be used whenever possible. Thus finally, CIE selected illuminant A and D65 as its standard illuminants, CIE standard illuminant A and CIE standard illuminant D65. The SPD of all illuminants defined by the CIE is available in the Technical Report Colorimetry [4] and the standard on CIE Standard Illuminants for Colorimetry [5].

CIE Standard Illuminant A

Originally CIE standard illuminant A (CIE St. Ill. A) was intended to represent typical, domestic, tungsten-filament lighting. Despite the fact that tungsten incandescent lamp light loses on importance, CIE St. Ill. A is the primary standard for calibrating photometers and colorimeters.

The relative spectral power distribution (SPD) of CIE St. Ill. A, $S_A(\lambda)$, is defined by the equation

$$S_A(\lambda) = 100 \left(\frac{560}{\lambda} \right)^5 \times \frac{\exp \frac{1,435 \times 10^7}{2,848 \times 560} - 1}{\exp \frac{1,435 \times 10^7}{2,848 \lambda} - 1}$$

where λ is the wavelength in nanometers. For practical applications, it is defined also in tabulated form over the wavelength range between 300 nm and 830 nm to six significant digits at 1 nm intervals [5]. The wavelength is to be taken as being in standard air (dry air at 15 °C and 101,325 Pa, containing 0.03% by volume of carbon dioxide). The numerical values in the two exponential terms are defined constants originating from the first definition of illuminant A [5].

CIE Standard Illuminant D65

The most important light source is daylight, the SPD of which changes during the day and depends also on weather conditions, etc.. From detailed measurements [7], the SPD of an often

encountered phase of daylight has been selected as primary reference spectrum. This SPD has a correlated color temperature of approximately 6,504 K and has been termed D65. CIE standardized this SPD as CIE standard illuminant D65 (CIE St. III. D65, or shortly D65); it should represent average daylight SPD. (For further details, see chapter on “Daylight Illuminants.”) For the time being, the relative SPD of CIE St. III. D65 is defined in tabulated form between 300 nm and 830 nm at 1 nm intervals. The wavelength values given apply in standard air. Intermediate values may be derived by linear interpolation.

CIE stated that CIE St. III. D65 should be used in all colorimetric calculations requiring representative daylight, unless there are specific reasons for using a different illuminant.

Short History

CIE Standard Illuminant A

CIE St. III. A was originally defined in 1931 as the relative spectral power distribution of a Planckian radiator of temperature $T = 2,848$ K, where Planck's equation has the form

$$M_{e,\lambda}(\lambda, T) = c_1 \lambda^{-5} [\exp(c_2/\lambda T) - 1]^{-1},$$

where $M_{e,\lambda}$ is radiant exitance (quotient of the radiant flux, $d\Phi_e$, leaving an element of the surface containing the point, by the area, dA , of that element).

In this equation, c_1 is not relevant, as $S_A(\lambda)$ is defined as a relative SPD; the c_2 constant's value was at the time of defining CIE St. III. A 14,350 $\mu\text{m}\cdot\text{K}$. The SPD of CIE St. III. A has not changed since this original definition, only the numerical value of c_2 has been several times reassigned in the International Temperature Scale (ITS), and due to this, the color temperature associated with CIE St. III. A has changed. For the different ITSs, the relevant color temperature values are

$$T_{27} = 2,842 \text{ K}, T_{48} = 2,854 \text{ K},$$

$$T_{68} = T_{90} = 2,856 \text{ K}.$$

Thus, the value we use nowadays is 2,856 K.

Daylight Illuminants

In 1931 CIE defined three reference illuminants, besides CIE St. III. A, also a spectrum that is near to that of direct sunlight, termed illuminant B (but this has been dropped after a short time) and illuminant C, with an SPD in the visible part of the spectrum representing average daylight with a correlated color temperature of approximately 6,774 K [6].

With the introduction of optical brighteners, it became necessary to define the SPD also in the near ultraviolet. The CIE accepted a recommendation by Judd and coworkers [7] to describe phases of daylight [8]. These authors found that although daylight is highly variable, the chromaticities of different phases of daylight fall on a curve more or less parallel to the Planckian locus on the chromaticity diagram. It turned out that the SPDs of the different phases of daylight can be described using only three basic functions, termed $S_0(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$; see Fig. 1.

The SPD of a phase of daylight can be calculated using the following equation:

$$S(\lambda) = S_0(\lambda) + M_1 S_1(\lambda) + M_2 S_2(\lambda)$$

where the M_1 and M_2 factors are calculated by the help of the chromaticity coordinates x_D , y_D of the phase of daylight:

$$M_1 = \frac{-1.3515 - 1.7703x_D + 5.9114y_D}{0.0241 + 0.2562x_D - 0.7341y_D},$$

$$M_2 = \frac{0.0300 - 31.4424x_D + 30.0717y_D}{0.0241 + 0.2562x_D - 0.7341y_D}.$$

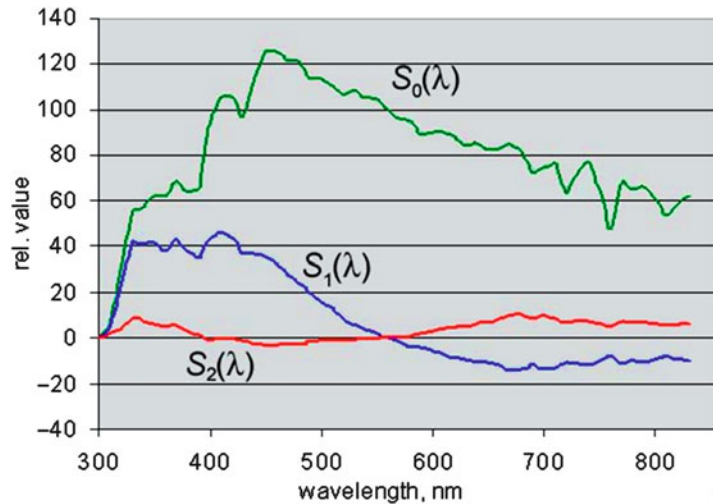
x_D , y_D are defined by the help of the correlated color temperature (T_{cp}) of the phase of daylight. The x_D coordinate has been defined in two parts for temperatures between 4,000 K and 7,000 K as

$$x_D = \frac{-4.6070 \times 10^9}{(T_{cp})^3} + \frac{2.9678 \times 10^6}{(T_{cp})^2} + \frac{0.09911 \times 10^3}{(T_{cp})} + 0.244063$$

and between 7,000 K and 25,000 K as

CIE Standard Illuminants and Sources,

Fig. 1 Characteristic vectors used to reconstitute phases of daylight



$$x_D = \frac{-2.0064 \times 10^9}{(T_{cp})^3} + \frac{1.9018 \times 10^6}{(T_{cp})^2} + \frac{0.24748 \times 10^3}{(T_{cp})} + 0.237040.$$

With the help of x_D , the corresponding y_D can be calculated as

$$y_D = -3.000 x_D^2 + 2.870 x_D - 0.275.$$

Using above equations, SPDs for any phase of daylight that has a T_{cp} between 4,000 K and 25,000 K can be calculated. From these spectra that corresponding to a T_{cp} of 6,500 was selected as primary standard: D65. As discussed already in connection with CIE St. Ill. A, the International Temperature Scale was based on a different value of the c_2 constant as used today. Thus, to keep the SPD of the D65 unchanged, the correlated color temperature had to be changed. To get to the correct spectrum, one has to insert in above equations the value of 6,504 instead of 6,500. (To be quite precise, the value should be 6503,616134, but for all practical purposes, 6,504 is accurate enough.) Equations have been elaborated that will produce the correct spectra for the rounded value of 6,500 K, and for that, the term “nominal correlated color temperature” is used.

Often SPDs are needed with slightly lower or higher T_{cp} s, thus, e.g., in graphic arts to get daylight spectra nearer to spectra used in indoor applications, a phase of daylight of $T_{cp} = 5,003$ K is used, termed D5000. Besides this, CIE publications often refer to D55 and D75 illuminants as well with T_{cp} of 5,503 and 7,504, respectively.

As can be seen from Fig. 1, the S functions are not smooth functions. Originally their values have been determined at 10 nm intervals and linear interpolation was suggested between these fixed points. A recent CIE recommendation deals with two possible smoothing algorithms, one that keeps the value of the SPD at the fixed points constant [9] and one that smoothes the curves further, so that they can be realized more accurately with physical sources [10, 11]. For further details, see entry on ► “Daylight Illuminants”.

Further CIE Illuminants

Often it is necessary to perform a colorimetric calculation using light source spectra of commercial lamps. To make such calculations more transparent, CIE published the spectra of some commercial gas-discharge lamps [4]: The collection of fluorescent lamp spectra includes 12 older constructions (six halophosphate, three broadband, and three narrowband lamp spectra); a second series contains, besides halophosphate lamps

(three spectra), three Deluxe lamp spectra, five three-band lamp spectra, three multiband spectra, and a D65 simulator spectrum. Further spectra show typical high-pressure sodium lamp and high-pressure metal halide lamp spectra.

CIE Standard Source

The spectrum of coiled coil tungsten incandescent lamps is very near to the SPD of a black-body radiator; thus, CIE standard illuminant A can be realized by a gas-filled tungsten-filament lamp operating at a correlated color temperature of 2,856 K ($c_2 = 14,388 \times 10^{-2} \text{ m} \cdot \text{K}$). If the source is also to be used in the UV region, a lamp having an envelope or window made of fused quartz or silica must be used because glass absorbs the UV component of the radiation from the filament.

The spectrum of CIE standard illuminant D65 is too complicated to be realized with small enough error to represent a standard D65 source. Real sources to be used in visual observations as daylight sources are called “simulators,” and readers are directed to the entry ► “[CIE Method of Assessing Daylight Simulators](#)” for further details.

Cross-References

- [CIE Method of Assessing Daylight Simulators](#)
- [Daylight Illuminants](#)

References

1. CIE: ILV: International Lighting Vocabulary. CIE S 017/E:2011. (2011)
2. CIE: Colorimétrie, resolutions 1-4. Recueil des travaux et compte rendu des séances, Hutième Session Cambridge – Septembre 1931, pp. 19–29. Bureau Central de la Commission, Paris. The National Physical Laboratory Teddington, Cambridge at the University Press Cambridge, UK (1932)
3. CIE: Official recommendations, committee E-1.3.1 – colorimetry. In: Proceedings of the 15th Session, Vienna, 1963, vol. A, p. 35. CIE Publication 11 A (1964)
4. CIE Technical Report: Colorimetry, 3rd edn. Publication 15:2004, CIE Central Bureau, Vienna (2004)

5. Joint CIE/ISO Standard: Colorimetry – Part 2: CIE Standard Illuminants for Colorimetry. CIE S 014-2/E:2006/ISO 11664-2:2007(E)
6. Wyszecki, G., Stiles, W.S.: Color Science, Concepts and Methods, Quantitative Data and Formulae, 2nd edn, p. 145. Wiley, New York (1982)
7. Judd, D.B., MacAdam, D.L.: Wyszecki G (with the collaboration of Budde HW, Condit HR, Henderson ST, Simonds JL): Spectral distribution of typical daylight as function of correlated color temperature. J. Opt. Soc. Am. **54**, 1031–1040 (1964)
8. CIE: Recommendations on standard illuminants for colorimetry. In: Proceedings of the CIE Washington Session, vol. A, pp. 95–97, CIE Publ. 14. (1967), Vienna.
9. Kránicz, B., Schanda, J.: Re-evaluation of daylight spectral distributions. Color. Res. Appl. **25**(4), 250–259 (2000)
10. CIE: Methods for re-defining CIE D illuminants. CIE **204**, 2013 (2013)
11. Kosztyán, Z.S., Schanda, J.: Smoothing spectral power distribution of daylight. Color. Res. Appl. (2012). <https://doi.org/10.1002/col.21732>. Article first published online on 9 May 2012

CIE Standard Source A

- [CIE Standard Illuminants and Sources](#)

CIE Tristimulus Values

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Synonyms

[XYZ](#)

Definition

Tristimulus values are defined as the amounts of the 3 reference color stimuli, in a given trichromatic system, required to match the color of the stimulus considered.

Background

The International Commission on Illumination (CIE) recommended CIE color specification system as the basis for colorimetry [1]. The ISO and CIE are also provided a series of joint standards [2–4]. The fundamental of colorimetry is Tristimulus Values [4]. They are used for color communication and reproduction. The three key elements of color perception are: human vision system, light, and object. If missing one of three elements, we will not be able to perceive color. These elements have been defined by CIE as color matching functions of $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, or $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, and $\bar{z}_{10}(\lambda)$ for 2° and 10° fields of viewing respectively, across visual spectrum to represent the human population having normal color vision. They also known as CIE 1931 and 1964 standard colorimetric observers (see [2]). CIE also standardized some illuminants in terms of Spectral Power Distribution (SPD) ($S(\lambda)$) [3]. Additionally, CIE specified the illuminating and viewing geometry for measuring a reflecting surface. Each surface is defined by spectral reflectance, $R(\lambda)$, a ratio of the reflected light from a sample to that reflected by a perfect diffuser identically illuminated [1]. The typical instrument for measuring spectral reflectance is a spectrophotometer.

Thus any color can be specified by a triad of numbers called Tristimulus Values, or XYZ values (see Eq. 1). The details can be found in references [1, 4]. An easier explanation can be a color is matched by the amounts of standard red, green, and blue lights by a normal color vision observer under a particular standard illuminant. These are the integration of the products of the functions in three components over the visible spectrum, say 380 to 700 nm.

$$\begin{aligned} X &= k \int_{\lambda} S(\lambda) R(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= k \int_{\lambda} S(\lambda) R(\lambda) \bar{y}(\lambda) d\lambda \\ Z &= k \int_{\lambda} S(\lambda) R(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (1)$$

where k constant was chosen so that $Y = 100$ for the perfect reflecting diffuser. If the CIE1964 standard colorimetric observer is used in Eqs. 1 and 2 all terms except k , $S(\lambda)$ and $R(\lambda)$ should include a subscript of 10.

For measuring self-luminous colors such as color displays, TV, light sources, Eq. 2 should be used instead of Eq. 1. This is due to the fact that the object and illuminant are not defined. The P function represents the spectral radiance or spectral irradiance of the target stimulus. The areas of colors considered in display applications usually have quite small angular subtense and the CIE 1931 standard colorimetric observer is the appropriate one to use. The typical instrument for measuring spectral radiance or irradiance is a spectroradiometer.

$$\begin{aligned} X &= k \int_{\lambda} P(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= k \int_{\lambda} P(\lambda) \bar{y}(\lambda) d\lambda \\ Z &= k \int_{\lambda} P(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (2)$$

where k constant is chosen so that $Y = 100$ for the appropriate reference white.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [Daylight Illuminants](#)
- [Spectroradiometer](#)

References

1. CIE Pub. No. 15:2004: Colorimetry. Central Bureau of the CIE, Vienna
2. ISO 11664-1:2007(E)/CIE S 014-1/E:2006 Joint ISO/-CIE Standard: Colorimetry — Part 1: CIE Standard Colorimetric Observers. Central Bureau of the CIE, Vienna
3. ISO 11664-2:2007(E)/CIE S 014-2/E:2006 Joint ISO/-CIE Standard: Colorimetry — Part 2: CIE Standard

Illuminants for Colorimetry. Central Bureau of the CIE, Vienna

4. ISO 11664-3:2012(E)/CIE S 014-3/E:2012 Joint ISO/-CIE Standard: Colorimetry — Part 3: CIE Tristimulus Values. Central Bureau of the CIE, Vienna

$$u' = \frac{4x}{-2x + 12y + 3} \quad (3)$$

$$v' = \frac{9y}{-2x + 12y + 3} \quad (4)$$

where

$$x = \frac{X}{X + Y + Z} \quad \text{and} \quad y = \frac{Y}{X + Y + Z}. \quad (5)$$

CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space

János Schanda
Veszprém, Hungary

Synonyms

CIE 1976 $L^*u^*v^*$ color space; UCS diagram; Uniform chromaticity scale diagram

Definitions

Uniform Chromaticity Scale Diagram (UCS Diagram)

The CIE 1976 uniform chromaticity scale diagram is a projective transformation of the CIE x , y chromaticity diagram yielding perceptually more uniform color spacing (see CIE colorimetry standard [1]). It is produced by plotting, as abscissa and ordinate, respectively, quantities defined by the equations:

$$u' = \frac{4X}{X + 15Y + 3Z} \quad (1)$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \quad (2)$$

where X , Y , Z are the tristimulus values of the test color stimulus based on the CIE 1931 standard colorimetric system defined in ISO 11664-1/CIE S 014-1 [2].

Equivalent definition:

Uniform Color Space

The CIE 1976 $L^*u^*v^*$ color space is a three-dimensional, approximately uniform color space produced by plotting in rectangular coordinates, L^* , u^* , v^* , quantities defined by the equations

$$L^* = 116 f\left(\frac{Y}{Y_n}\right) - 116 \quad (6)$$

$$u^* = 13 L^* (u' - u'_n) \quad (7)$$

$$v^* = 13 L^* (v' - v'_n) \quad (8)$$

where

$$f\left(\frac{Y}{Y_n}\right) = \left(\frac{Y}{Y_n}\right)^{1/3} \quad \text{if} \quad \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \quad (9)$$

$$f\left(\frac{Y}{Y_n}\right) = \frac{841}{108} \left(\frac{Y}{Y_n}\right) + \frac{4}{29} \quad \text{if} \quad \frac{Y}{Y_n} \leq \left(\frac{6}{29}\right)^3. \quad (10)$$

In these equations, Y , u' , and v' describe the test color stimulus, and Y_n , u'_n , and v'_n describe a specified white stimulus.

If the angle subtended at the eye by the test stimulus is between about 1° and 4° , the tristimulus values X , Y , Z calculated using the color-matching functions of the CIE 1931 standard colorimetric system should be used. If this angular subtense is greater than 4° , the tristimulus values X_{10} , Y_{10} , Z_{10} calculated using the color-matching functions of the CIE 1964 standard colorimetric system should be used. The same color-matching functions and the same specified white stimulus should be used for all stimuli to be compared with each other.

Correlates of Lightness, Saturation, Chroma, and Hue

Approximate correlates of the perceived attributes lightness, saturation, chroma, and hue are calculated as follows:

CIE 1976 lightness	L^* as defined by Eq. 6
CIE 1976 u, v saturation (CIELUV saturation)	$s_{uv} = 13 \left[(u' - u'_n)^2 + (v' - v'_n)^2 \right]^{1/2}$
CIE 1976 u, v chroma (CIELUV chroma)	$C_{uv}^* = \left[(u^*)^2 + (v^*)^2 \right]^{1/2}$
CIE 1976 u, v hue angle (CIELUV hue angle)	$h_{uv} = \arctan (v^*/u^*)$

where the CIELUV hue angle lies between 0° and 90° if u^* and v^* are both positive, between 90° and 180° if v^* is positive and u^* is negative, between 180° and 270° if v^* and u^* are both negative, and between 270° and 360° if v^* is negative and u^* is positive.

Color Differences

The CIE 1976 $L^*u^*v^*$ color difference, ΔE_{uv}^* , between two color stimuli is calculated as the Euclidean distance between the points representing them in the space:

$$\Delta E_{uv}^* = \left[(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2 \right]^{1/2}$$

$$\text{or } \Delta E_{uv}^* = \left[(\Delta L^*)^2 + (\Delta C_{uv}^*)^2 + (\Delta H_{uv}^*)^2 \right]^{1/2} \quad (11)$$

$$\text{where } \Delta H_{uv}^* = 2 \left(C_{uv,1}^* C_{uv,2}^* \right)^{1/2} \sin(\Delta h_{uv}/2).$$

For further details and calculating differences of color coordinate components, see Eq. 1.

Overview

The CIE 1976 uniform chromaticity scale diagram and the CIE 1976 $L^*u^*v^*$ color space have

been agreed by the CIE in 1976 as a modification of the CIE1960 UCS diagram (u, v diagram) and CIE 1964 uniform color space (U^*, V^*, W^* space). The u, v diagram was devised by David MacAdam [3] and recommended by the CIE in 1959 [4]. Observations have shown that the u, v diagram could be made more uniform if v coordinate would be modified. The original equations for u, v are given in Eq. 12.

$$u = 4X/(X + 15Y + 3Z) \text{ and}$$

$$v = 6Y/(X + 15Y + 3Z);$$

thus, $u' = u$, and $v' = 3v/2$ [5].

Figure 1 shows MacAdam ellipses in the u', v' diagram. As can be seen, the u', v' diagram is not absolutely equidistant (the ellipses are not circles), but its uniformity is by an order of magnitude better as that of the x, y diagram.

The u, v diagram described only chromaticities of colors of equal luminance. In 1963, CIE extended this to a three-dimensional spacing perceptually more nearly uniform than that provided by the (XYZ) system. “The recommended coordinate system is formed by plotting the variables U, V , and W along orthogonal axes where U, V , and W are defined in terms of the tristimulus values X, Y, Z , as

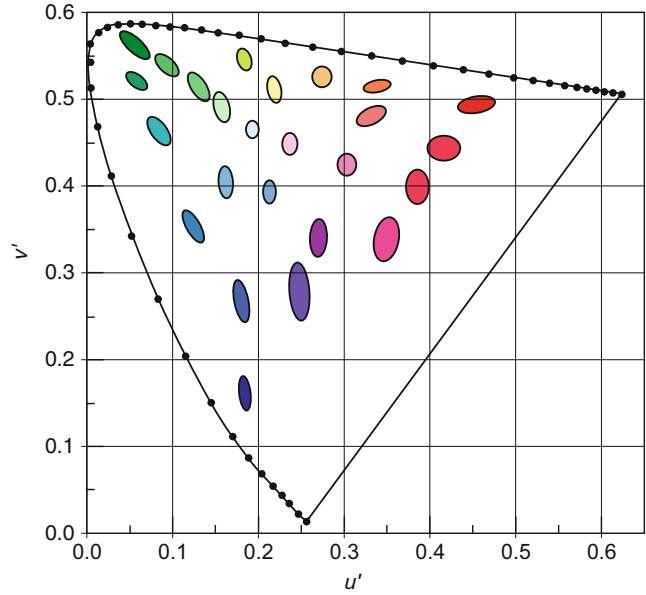
$$W = 25Y^{1/3} - 17, 1 \leq Y \leq 100;$$

$$U = 13W(u - u_n); \quad V = 13W(v - v_n)$$

where u, v, u_n , and v_n are described in the previous paragraphs [6].

The establishment of the CIE uniform chromaticity scale diagram and the CIELUV space was the result of a long-lasting development. The experts, at the CIE 1976 meeting, proposed two systems, the CIELUV and the CIELAB systems, but on the then available visual data, they were unable to decide between the two. During the past 30–40 years, the CIELAB system and its derivatives (see CIELAB in this volume) proved to be a better approximation of color perception; nowadays, the CIELUV space has only very limited use, but the u', v' diagram is still in general use, as it represents a linear diagram, in which chromaticities can easily be calculated. Thus, e.g., if the

CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space, Fig. 1 The CIE u' , v' diagram with ten times enlarged MacAdam ellipses (ellipses showing the just noticeable color difference); colors show approximate chromaticity regions in the u' , v' diagram. (From Schubert Light emitting diodes, Chapter 17 Colorimetry, Fig. 17.7, with the kind permission of Cambridge Univ. Pr)



chromaticity of a blue LED and of a phosphor emission that is excited by the LED is given, it is easy to visualize the chromaticities that can be produced by the additive mixture of the two colors; they are located on the straight line drawn between the two chromaticity points.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Tristimulus Values](#)
- [CIELAB](#)

References

1. IE: Joint International Standard: Colorimetry – Part 5: CIE 1976 $L^*u^*v^*$ colour space and u' , v' uniform chromaticity scale diagram. CIE S 014-5/E:2009/ISO 11664-5 (2009)
2. CIE: Joint International Standard: Colorimetry – Part 1: CIE Standard Colorimetric Observers. CIE S 014-2/E:2006/ISO 11664-2:2007(E)
3. MacAdam: Projective transformations of I. C. I. color specifications. JOSA. **27**, 294–299 (1937)

4. CIE: Official Recommendation of the Committee W-1.3.1 – Colorimetry on “a diagram yielding colour spacing perceptually more nearly uniform than the (x y) diagram.” In: Proceedings of the Brussels Session 1959. CIE Publication 4–7, vol. A, pp. 3–37 & 91–94. CIE (1960)
5. CIE: Progress report of CIE TC – 1.3 Colorimetry. Comptes Rendus 18e Session, Londres 1975. CIE Publication 36, pp. 161–172. (1975)
6. CIE: Official Recommendations of the Committee E-1.3.1 – Colorimetry. In: Proceedings of the Vienna Session 1963. CIE Publication 11. vol. A, pp. 35 & 112–113. CIE (1963)

CIE Whiteness

Stephen Westland
Colour Science and Technology, University of
Leeds, Leeds, UK

Definition

The Commission Internationale de l'éclairage (CIE) whiteness formula was recommended in 1986 as an assessment method for white materials [1]. For the CIE 1931 standard colorimetric observer, the whiteness index W is given by

$$W = Y + 800 (x_n - x) + 1,700 (y_n - y), \quad (1)$$

where x, y are the chromaticity coordinates of the sample, and x_n, y_n are those of the illuminant. For the CIE 1964 supplementary standard colorimetric observer, the whiteness index W_{10} is given by

$$W_{10} = Y + 800 (x_{n,10} - x_{10}) + 1,700 (y_{n,10} - y_{10}), \quad (2)$$

where x_{10}, y_{10} and $x_{n,10}, y_{n,10}$ are the chromaticity coordinates of the sample and the illuminant, respectively.

The tint coefficient T_w or $T_{w,10}$, given by the following formulae, is zero for a sample without reddishness or greenishness. For the CIE 1931 standard colorimetric observer,

$$T_w = 1,000 (x_n - x) - 650 (y_n - y), \quad (3)$$

and for the CIE 1964 supplementary standard colorimetric observer,

$$T_{w,10} = 900 (x_{n,10} - x_{10}) - 650 (y_{n,10} - y_{10}), \quad (4)$$

where x_{10}, y_{10} and $x_{n,10}, y_{n,10}$ are the chromaticity coordinates of the sample and the illuminant, respectively. The tint coefficient allows two samples with the sample whiteness index, W or W_{10} , but with different hue to be distinguished. The more positive the value of T_w or $T_{w,10}$, the greater is the indicated greenishness, and the more negative, the greater the reddishness [2].

The higher the value of W or W_{10} , the higher the whiteness of the sample. However, the CIE standard includes the following restriction:

The application of the formulae is restricted to samples that are called 'white' commercially, that do not differ much in color and fluorescence, and that are measured on the same instrument at nearly the same time; within these restrictions, the formulae provide relative, but not absolute, evaluations of whiteness that are adequate for commercial use, when employing measuring instruments having suitable modern and commercially available facilities. The following

boundaries are proposed for the application of whiteness formulae:

$$40 < W \text{ or } W_{10} < 5Y - 280; \quad -3 < T_w \text{ or } T_{w,10} < +3, \quad (5)$$

where Y is the tristimulus value of the sample.

Development of the CIE Whiteness Formula

Whiteness is a commercially important property of color appearance in a number of industries, most notably those of textiles and paper. According to Wyszecki and Stiles, white is the attribute of a visual sensation according to which a given stimulus appears to be void of any hue and grayness [3]. Alternatively, the percept of whiteness is caused by a combination of high lightness and lack of yellowness [4]. MacAdam published the first instrumental method for the assessment of whiteness in 1934 [5], and by the 1960s, over 100 different methods to predict perceptual whiteness had been proposed. Although color is a three-dimensional percept, whiteness has traditionally been measured using a univariant metric. Hayhurst and Smith suggest that the reason for this is that the color of white textile or paper is often influenced by the quantity of a specific (single) impurity in it [4]; the white color therefore falls on a line in three-dimensional color space, albeit one that is curved. However, especially when different products are being compared, it is important to have a metric that is at least dependent upon all three dimensions of color space even if the metric is condensed to a single number, as is the case with the CIE whiteness method. Many of the early whiteness formulae considered only one or two of the three dimensions of color and therefore had limited general applicability.

The foundations for a modern and effective whiteness formula began with Ganz [6] who proposed a new generic formula for whiteness W_G in 1972 with the following form:

$$W_G = DY + Px + Qy + C, \quad (6)$$

where Y and x , y are the Y tristimulus value (sometimes called luminous reflectance) and chromaticity coordinates, respectively, of the sample, and D , P , Q , and C are coefficients that could be varied to adjust for various parameters including the illuminant, the observer, and the hue preference. The coefficient of D was set to unity and $C = P_{x,n} + Q_{y,n}$. Ganz suggested three pairs of values for P and Q each corresponding to a different hue preference. The formula was linear in chromaticity space because experience in producing white scales had shown that the chromaticities of uniformly spaced samples turn out to be approximately equidistant in this space [7].

A round-robin test [7], organized by a technical committee (TC-1.3) of the CIE in the 1970s, showed that there are individuals with extreme hue preference for whiteness. This suggests that a single whiteness formula cannot reproduce the whiteness preferences of a sizable portion of the population and this is why Ganz originally suggested three pairs of values for P and Q . The CIE adopted a form of Eq. 6 with values of P and Q corresponding to neutral hue preference, and the CIE equation is normally written in the forms

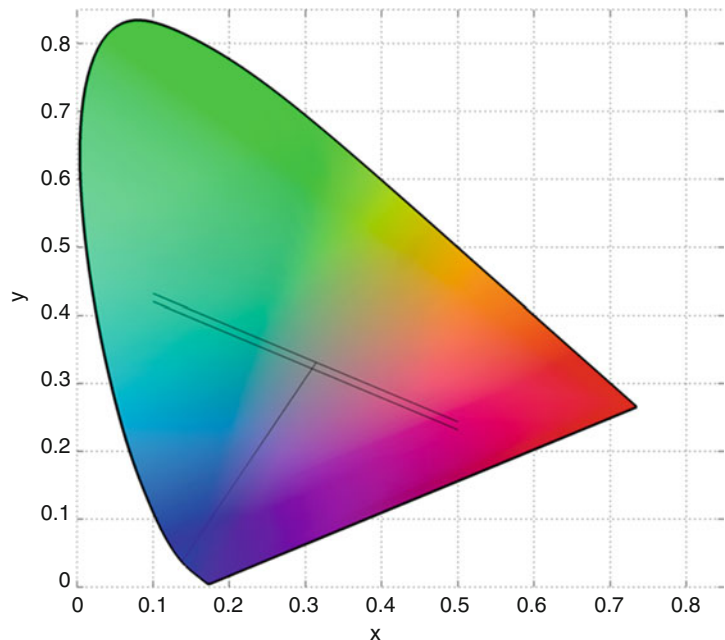
of Eqs. 1 and 2. The use of unity for the coefficient D has the implication that the whiteness of a perfect reflecting diffuser would be 100. Samples containing fluorescent whitening agents may have W or $W_{10} \gg 100$. The difference that is just perceptible to an experienced visual assessor is about three CIE whiteness units [4].

Ganz and Griesser developed a generic formula for the instrumental evaluation of tint [8]. Lines of equal tint run approximately parallel to the line of dominant wavelength 470 nm (467.6 and 464.7 nm, respectively, for the 1931 and 1964 CIE standard observers) in the chromaticity diagram. Figure 1 shows the line of dominant wavelength 470 nm and also two lines of iso-whiteness according to the CIE whiteness equation. The CIE adopted the Ganz-Griesser tint formula as part of the CIE whiteness standard.

The whiteness formula should be used only for the comparison of samples that are commercially white and do not differ much in color or fluorescence. The samples should be measured on the same instrument at the same time. The light source in the spectrophotometer used to measure the reflectance factors should match, as closely as possible, the illuminant D65. The requirement for instruments that could deliver consistent light

CIE Whiteness,

Fig. 1 CIE 1964 chromaticity diagram showing the line at dominant wavelength of 465 nm and two lines of iso-whiteness. The *upper* iso-whiteness line shows the locus of points with $W_{10} = 100$, passing through the white point; each point on this line has the same whiteness but a different tint. The *lower* iso-whiteness line shows the locus of points with $W_{10} = 120$. Whiteness generally increases from the *yellowish* part of the diagram to the *bluish* part



output (including in the near UV) that closely matches illuminant D65 led to developments in instrument-calibration techniques toward the end of the twentieth century.

Further Considerations and Future Directions

Since the introduction of the CIE whiteness equation, the approximately uniform color space, CIELAB, has been developed. This led some practitioners to want to calculate CIE whiteness in CIELAB space directly. Ganz and Pauli [9] developed the following equations that approximate the CIE whiteness and tint equations in the CIE (1976) $L^*a^*b^*$ space for the 1964 standard observer:

$$W_{10} = 2.41L^* - 4.45b^*[1 - 0.0090(L^* - 96)] - 141.4, \quad (7)$$

and

$$T_{w,10} = -1.58a^* - 0.38b^*, \quad (8)$$

where L^* , a^* , and b^* are the CIELAB coordinates of the sample. The upper limit $W_{10} < 5Y - 280$ was transformed by a linear approximation yielding $W_{10} < 10.6L^* - 852$. The other limits $W_{10} > 40$ and $-3 < T_{w,10} < 3$ remained unchanged.

The CIE whiteness formula has been found to correlate with visual assessments for many white samples where the tint is similar and where the level of fluorescence is comparable [10]. However, the level of agreement is much less when there are different tints and levels of fluorescence [10]. Observers frequently give the highest whiteness estimation to fluorescent-whitened samples with a hint of red, blue, or green. This is called “preferred white.” Uchida considered data from 26 observers who evaluated the whiteness of 49 fluorescent-whitened samples and devised a new whiteness formula W_u where

$$W_u = W_{10} - 2(T_{w,10})^2 \quad (9)$$

The Uchida equation [10] was recommended for samples where $40 < W_{10} < 5Y - 275$ and was found to give better agreement with the visual results than the CIE equation. In a recent study [11], 22 observers were asked to rank 20 samples (with low to high CIE whiteness indices). Results showed a significant consistency between the variations in the ordering decisions of the observers for the white samples with low CIE whiteness index but a high level of disagreement between the observers for the whiter samples.

Recently the generic form of the CIE whiteness equation (Eq. 2) with modified coefficients has been shown to be able to predict the whiteness of human teeth even though many of the samples would not be considered to be commercially white in the textile, plastic, and paper industries in which the CIE whiteness equation has traditionally been applied [12].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Blackbody and Blackbody Radiation](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)

References

1. CIE, Colorimetry Second Edition: The Evaluation of Whiteness, pp. 36–38. CIE Publication 15.2 (1986)
2. Hunt, R.W.G.: The Reproduction of Colour. Wiley, Chichester, UK (2004)
3. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae. Wiley, Hoboken, USA (1982)
4. Hayhurst, R., Smith, K.J.: Instrumental evaluation of whiteness. *J. Soc. Dye. Colour.* **111**, 263–266 (1995)
5. MacAdam, D.L.: The specification of whiteness. *J. Opt. Soc. Am.* **24**, 188–191 (1934)
6. Ganz, E.: Whiteness measurement. *J. Color Appearance.* **1**, 33–41 (1972)
7. Ganz, E.: Whiteness formulas: a selection. *Appl. Opt.* **18**(7), 1073–1078 (1979)
8. Ganz, E., Griesser, R.: Whiteness: assessment of tint. *Appl. Opt.* **20**(8), 1395–1396 (1981)
9. Ganz, E., Pauli, H.K.A.: Whiteness and tint formulas of the Commission Internationale de l’Eclairage: approximations in the $L^*a^*b^*$ color space. *Appl. Opt.* **34**(16), 2998–2999 (1995)

10. Uchida, H.: A new whiteness formula. *Color. Res. Appl.* **23**(4), 202–209 (1998)
11. Jafari, R., Amirshahi, S.H.: Variation in the decisions of observers. *Color. Technol.* **124**(2), 127–131 (2008)
12. Luo, W., Westland, S., Ellwood, R., Pretty, I., Cheung, V.: Development of a whiteness index for dentistry. *J. Dent.* **37**, e21–e26 (2009)

CIE94, History, Use, and Performance

Manuel Melgosa

Optics Department, University of Granada,
Granada, Spain

Synonyms

CIE 1994 (ΔL^* ΔC^*_{ab} ΔH^*_{ab}); ΔE^*_{94} ; Hue Difference, Delta H

Definition

The CIE94 color-difference formula [1] was developed by the CIE Technical Committee 1–29 “Industrial Color Difference Evaluation” chaired by Dr. D.H. Alman (USA). CIE94 is a CIELAB-based color-difference formula where CIELAB lightness, chroma, and hue differences are appropriately weighted by very simple “weighting functions” correcting CIELAB’s lack of visual uniformity, as well as by “parametric factors” accounting for the influence of illuminating/viewing conditions in color-difference evaluation. CIE94 may be considered a simplified version of the CMC color-difference formula [2], as well as a predecessor of currently CIE-recommended color-difference formula, the joint ISO/CIE standard CIEDE2000 [3, 4].

Overview

In 1976, CIE recommended the CIELUV and CIELAB color spaces with their corresponding color-difference formulas, defined as Euclidean distances in such spaces. These two color spaces

are only approximately uniform, and it was indicated that the use of different weighting for lightness, chroma, and hue differences may be necessary in different practical applications [5]. CIELAB has been widely accepted in many industrial applications [6], and in the past recent years, several CIELAB-based color-difference equations have been proposed to improve its performance. Among these formulas, we can mention, in chronological order, JPC79 [7], CMC [2], BFD [8], and CIE94. Nowadays, it is generally agreed that most of these formulas significantly improved CIELAB [9].

Assuming that subindices 1 and 2 indicate the two samples in a color pair, the next equations define the CIE94 color-difference formula, designated as $\Delta E^*_{94}(k_L:k_C:k_H)$:

$$\Delta E^*_{94}(k_L : k_C : k_H) = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H S_H}\right)^2} \quad (1)$$

where the CIELAB lightness, chroma, and hue differences are given by

$$\Delta L^* = L^*_1 - L^*_2 \quad (2)$$

$$\Delta C^*_{ab} = C^*_{ab,1} - C^*_{ab,2} \quad (3)$$

$$\Delta H^*_{ab} = 2\sqrt{C^*_{ab,1}C^*_{ab,2}} \sin\left(\frac{\Delta h_{ab}}{2}\right) \quad (4)$$

$$\Delta h_{ab} = h_{ab,1} - h_{ab,2}. \quad (5)$$

The “weighting functions” for lightness (S_L), chroma (S_C), and hue (S_H) are defined as

$$S_L = 1 \quad (6)$$

$$S_C = 1 + 0.045\left(C^*_{ab,1}C^*_{ab,2}\right)^{1/2} \quad (7)$$

$$S_H = 1 + 0.015\left(C^*_{ab,1}C^*_{ab,2}\right)^{1/2}, \quad (8)$$

and the “parametric factors” k_L , k_C , k_H are set as 1.0 (i.e., do not affect the total color difference computation) under the next illuminating/viewing

conditions, usually known as “reference conditions”:

- Illumination: D65 source
- Illuminance: 1000 lx
- Observer: Normal color vision
- Background field: Uniform, neutral gray with $L^* = 50$
- Viewing mode: Object
- Sample size: Greater than 4 degrees
- Sample separation: Direct edge contact
- Sample color-difference magnitude: Lower than 5.0 CIELAB units
- Sample structure: Homogeneous (without texture)

In the textile industry, it is common practice to set the lightness parametric factor k_L to 2. Although the experimental conditions leading to this parametric correction to lightness difference are not yet well understood, it introduces important improvements in the performance of CIE94, which in this case must be designated as ΔE^*_{94} (2:1:1) or CIE94 (2:1:1).

For a constant ΔE^*_{94} value, Eq. 1 approximately represents [10] an ellipsoid with semiaxis lengths given by the denominators $k_L S_L$, $k_C S_C$, $k_H S_H$. Under reference conditions, the semiaxis lengths S_L , S_C , S_H are often denominated lightness, chroma, and hue tolerances, respectively. Equation 6 indicates that lightness tolerance is the same for all color centers. However, Eq. 7

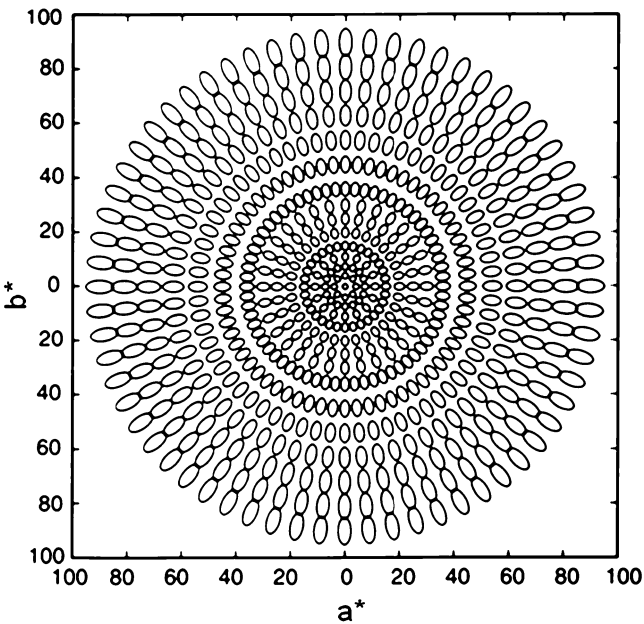
indicates that chroma tolerances increase with chroma; that is, human sensitivities to chroma differences are smaller for color centers with higher chroma, as earlier pointed out by McDonald [11]. Similarly, Eq. 8 also indicates that hue tolerances/sensitivities increase/decrease with chroma. In summary, assuming small color differences [10], the loci of constant CIE94 differences in CIELAB color space can be represented by ellipsoids with constant lightness semiaxes, which sections in the a^* , b^* plane are ellipses with their major axes pointing to the origin (i.e., ellipses radially oriented). The major and minor semiaxes of these a^* , b^* ellipses linearly increase with the chroma of the ellipse centers (Fig. 1).

Development

The development of CIE94 [12, 13] began with a selection of experimental visual datasets meeting the next main conditions: statistical significance (i.e., to represent a population average with its corresponding uncertainty), well-documented experimental conditions, and use of object color specimens. CIE TC 1–29 decided to use only three experimental datasets: Witt [14, 15], Luo and

CIE94, History, Use, and Performance,

Fig. 1 CIE94 color-tolerance ellipses (or contours of approximately constant CIE94 units) in the a^* , b^* plane. (Figure from R. S. Berns [13], p. 121, graph courtesy of S. Quan)



Rigg [16], and RIT-DuPont [17, 18]. The main goal was to use these datasets to find the best weighting functions S_L , S_C , S_H correcting CIELAB. The analyses also considered the characteristics of the previous CMC color-difference formula, which was an ISO standard in textiles [19] and was successfully employed in different industries.

As indicated by Eqs. 7 and 8, it was found that simple linear chroma functions described the main trends in the experimental datasets analyzed. Both the lightness dependence of lightness tolerances and the hue angle dependence of hue tolerances proposed by the CMC color-difference formula were found not robust trends in the experimental datasets analyzed [13], and therefore they were disregarded in CIE94. It is possible that these two corrections proposed by CMC were influenced by some specific parametric factors. Anyway, it can be said that CIE94 was a conservative approach incorporating only the main corrections to CIELAB. Currently, the CIEDE2000 color-difference formula has incorporated additional corrections to CIELAB. For example, CIEDE2000 proposes a V-shaped function for the S_L function, which is different to both the lightness function proposed by CMC and the simple $S_L = 1$ adopted by CIE94.

CIE94 was the first CIE-recommended color-difference formula incorporating the influence of illuminating/viewing experimental conditions in color-difference evaluations through the use of the so-called parametric factors k_L , k_C , k_H (see Eq. 1).

The work carried out by CIE TC 1–29 finished with the proposal of CIE94 and some guidelines for coordinated future work on industrial color-difference evaluation [20]. These guidelines updated those earlier given by Robertson [21], proposing a new set of 17 color centers to be studied in future research, considering the effects of changes from the “reference conditions,” and suggesting the development of a database of color-difference visual responses.

Performance

From the combined dataset employed at CIEDE2000 development [22], the performance

of different color-difference formulas has been tested using the *STRESS* index [23]. Low *STRESS* values (always in the range 0–100) indicate better color-difference formula performance. *STRESS* values for CIELAB, CIE94, and CIEDE2000 are 43.9, 32.1, and 27.5, respectively [24]. As we can see, the improvement achieved by CIE94 upon CIELAB was considerably higher than the one achieved by CIEDE2000 with respect to CIE94. From *STRESS* values, it can be also concluded that in CIE94 the S_C function (which was also adopted in CIEDE2000) is a much more important correction to CIELAB than the S_H function [24]. Besides CIE94 being proposed for object colors, it has been reported that this formula also performed satisfactorily for self-luminous color datasets [25].

Curiously, Eqs. 7 and 8 involve the geometrical mean of the CIELAB chroma of the two samples in the color pair, in place of the more simple arithmetical mean proposed in other color-difference formulas, for example, CIEDE2000. Strictly speaking, this implies that the locus of constant CIE94 differences with respect to a given color center is not an ellipsoid/ellipse, although deviations from ellipsoidal/elliptical contours can be considered negligible in most practical situations. CIE94 color differences computed using the geometrical and arithmetical means of the CIELAB chroma of the samples in the color pair are slightly different, in particular for colors with very low chroma [26].

In comparison with other recent formulas like JPC79, CMC, BFD, or CIEDE2000, it can be said that CIE94 was relevant because it was a very simple and versatile color-difference formula accounting for the main robust trends found in reliable color-difference visual datasets. CIE94 just proposes to use simple corrections to CIELAB provided by the weighting functions S_L , S_C , S_H plus consideration of the influence of the illuminating/viewing conditions using the parametric factors k_L , k_C , k_H . After the CIE94 proposal, CIE TC 1–47 continued further work leading to the last CIE-recommended color-difference formula, CIEDE2000, which significantly improved CIE94 for the experimental datasets used in its development [22].

Cross-References

- [CIEDE2000, History, Use, and Performance](#)
- [CIELAB](#)

References

1. CIE Publication 116: Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (1995)
2. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dye. Colour.* **100**, 128–132 (1984)
3. CIE Publication 142: Improvement to Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (2001)
4. CIE S 014-6/E:2013: Colorimetry – Part 6: CIEDE2000 Colour-Difference Formula. CIE Central Bureau, Vienna (2013)
5. CIE Publication 15.2: Colorimetry, 2nd edn, Note 9, p. 33. CIE Central Bureau, Vienna (1986)
6. Kuehni, R.G.: Industrial color-difference: progress and problems. *Color. Res. Appl.* **15**, 261–265 (1990)
7. McDonald, R.: Industrial pass/fail colour matching. Part III: development of a pass/fail formula for use with instrumental measurement of colour difference. *J. Soc. Dye. Colour.* **96**, 486–497 (1980)
8. Luo, M.R., Rigg, B.: BFD(l:c) colour-difference formula. Part 1 – development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
9. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
10. Brill, M.H.: Suggested modification of CMC formula for acceptability. *Color. Res. Appl.* **17**, 402–404 (1992)
11. McDonald, R.: The effect of non-uniformity in the ANLAB color space on the interpretation of visual colour differences. *J. Soc. Dye. Colour.* **90**, 189–198 (1974)
12. Berns, R. S.: The mathematical development of CIE TC 1–29 proposed color difference equation: CIELCH. In: Nemcsics, A., Schanda, J. (eds.) AIC (Association Internationale de la Couleur). 1993. AIC Color 93, Proceedings of the 7th Congress, Budapest, 13–18 June 1993, 3 vols. Hungarian National Colour Committee, Budapest. (1993)
13. Berns, R.S.: Billmeyer and Saltzman's Principles of Color Technology, pp. 120–121. Wiley, New York (2000)
14. Witt, K., Döring, G.: Parametric variations in threshold color-difference ellipsoid for green painted samples. *Color. Res. Appl.* **8**, 153–163 (1983)
15. Witt, K.: Three-dimensional threshold color-difference perceptibility in painted samples: variability of observers in four CIE color regions. *Color. Res. Appl.* **12**, 128–134 (1987)
16. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colors. *Color. Res. Appl.* **11**, 25–42 (1986)
17. Alman, D.H., Berns, R.S., Snyder, G.D., Larsen, W.A.: Performance testing of color-difference metrics using a color tolerance dataset. *Color. Res. Appl.* **14**, 139–151 (1989)
18. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of supra-threshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
19. ISO 105-J03: Textiles: Test for Colour Fastness – Part 3: Calculation of Colour Differences. ISO, Geneva (2009)
20. Witt, K.: CIE guidelines for coordinated future work on industrial colour-difference evaluation. *Color. Res. Appl.* **20**, 399–403 (1995)
21. Robertson, A.R.: CIE guidelines for coordinated research on color-difference evaluation. *Color. Res. Appl.* **3**, 149–151 (1978)
22. Luo, M.R., Cui, G., Rigg, B.: The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color. Res. Appl.* **26**, 340–350 (2001)
23. García, P.A., Huertas, R., Melgosa, M., Cui, G.: Measurement of the relationship between perceived and computed color differences. *J. Opt. Soc. Am. A.* **24**, 1823–1829 (2007)
24. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A.* **25**, 1828–1834 (2008)
25. Melgosa, M., Hita, E., Pérez, M.M., El Moraghi, A.: Sensitivity differences in chroma, hue, and lightness from several classical threshold datasets. *Color. Res. Appl.* **20**, 220–225 (1995)
26. Melgosa, M., Huertas, R., Hita, E., Benítez, J. M.: Diferencias de color CIE94 con/sin un estímulo de referencia. In: Proceedings VI Congreso Nacional del Color (ISBN 84–699–9187-6), Area de Nutrición y Bromatología, pp. 117–118. Universidad de Sevilla (2002)

CIECAM02

Nathan Moroney
Hewlett-Packard Laboratories, Palo Alto,
CA, USA

Synonyms

[CIE color appearance model 2002](#)

Definition

CIECAM02 is a color appearance model that provides a viewing condition specific method for transforming between tristimulus values and perceptual attribute correlates. This model was first published [1] in 2002 by Division 8 of the International Commission on Illumination (CIE). CIECAM02 was developed for use in color management systems and was based on the previously published CIECAM97s color appearance model [2, 3]. The model provides a number of parameters for defining a viewing condition and also inverse equations for transforming perceptual attribute correlates back to tristimulus values for a given set of viewing conditions. In this way, CIECAM02 can be used to transform perceptual attribute correlates, such as lightness, chroma, and hue, across different viewing conditions.

Overview

A color appearance model [4, 5] transforms between colorimetry, which specifies if stimuli match, and perceptual attribute correlates, which are scales of lightness, chroma, and hue. To do so,

CIECAM02 provides a set of viewing condition parameters in order to model specific color appearance phenomena, such as chromatic adaptation and simultaneous contrast. The resulting perceptual attribute correlates can then be used in research and engineering applications requiring a viewing condition independent color representation, such as color calibration of color printers.

The viewing conditions for CIECAM02 consist of the background, the adapting field, the surround, the white point, and the luminance of the adapting field. Figure 1 shows an example of a color stimulus, a background, and an adapting field. Note that this example is of a real-world situation of viewing a reflectance print. The central stimulus is a 2° region that corresponds to a portion of the print the size of a thumb viewed at arm's length. The background is the 10° region surrounding the stimulus and is roughly the size of a fist viewed at arm's length. The adapting field is everything else in the field of view.

The luminance of the adapting field can be measured directly with an illuminance meter. To incorporate a gray-world assumption and convert to luminance, this is then divided by 5π . The surround setting for the model is categorical and follows roughly the specific application. Dark



CIECAM02, Fig. 1 An example of a colored stimulus, background, and adapting field. The stimulus subtends 2° or roughly the size of a thumb viewed at arm's length. The background subtends 10° or approximately the size of a fist viewed at arm's length. The adapting field is everything

else in the field of view. The stimulus is shown as the *central smaller black circle*. The background is shown as the *larger black circle*. The adapting field is everything outside of the *larger black circle*

surrounds are those with no ambient illumination or viewing film projected in a darkened room. Dim surrounds are those in which the ambient illumination is not zero but is also less than 20% of the scene, print, or display white point, such as home viewing of television with low light levels. Average surround is ambient illumination greater than 20% of the scene, print, or display white point, such as viewing of surface colors in a light booth. The CIECAM02 model has a set of constants associated with each surround.

Finally the model has an associated white for all calculations. The selection of a white point is a subtle topic and the model suggests two approaches to this issue. First is to use an adopted white point or a computational white point for all calculations. An adopted white point is a fixed value, such as one based on a standard viewing condition or to ensure a specific final mapping for the white point. Second is an adapted white or the white point adapted by the human visual system for a given set of viewing conditions. An adapted white point is one which attempts to as closely as possible match the state of adaptation for a human observer. Note that adapted white points may require experimentation to infer their value, such as for a novel set of viewing conditions or situations with multiple illuminants. In cases where it is not possible to determine the adapted white

point, use of an adopted or assumed white point can be convenient.

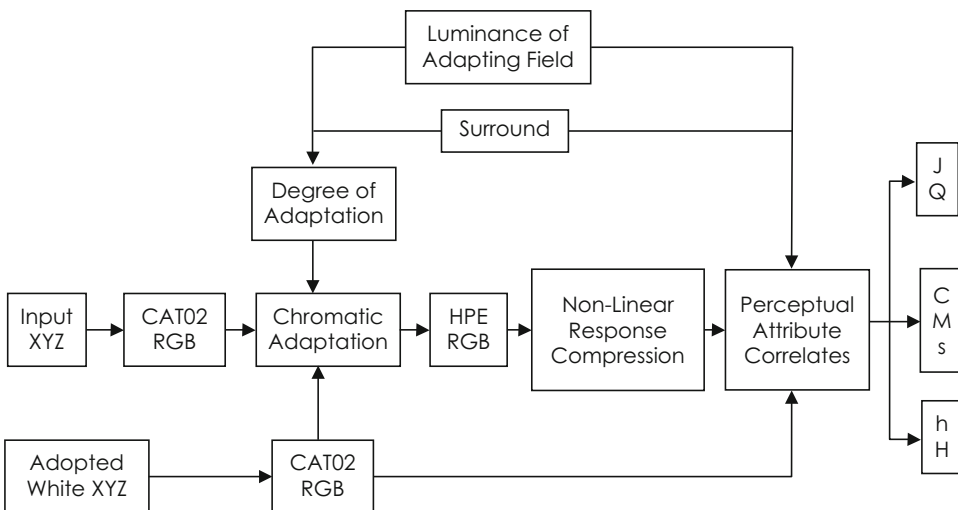
Given a set of viewing conditions and their associated parameters, it is then possible to compute the perceptual attribute correlates. Figure 2 shows an overall flowchart of the CIECAM02 model starting with input tristimulus values and adopted white point tristimulus values on the left.

The first step is to apply a matrix to convert the XYZ values to the CAT02 RGB space. The space was selected as a preferred color space to perform chromatic adaptation. This matrix can be written:

$$M_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (1)$$

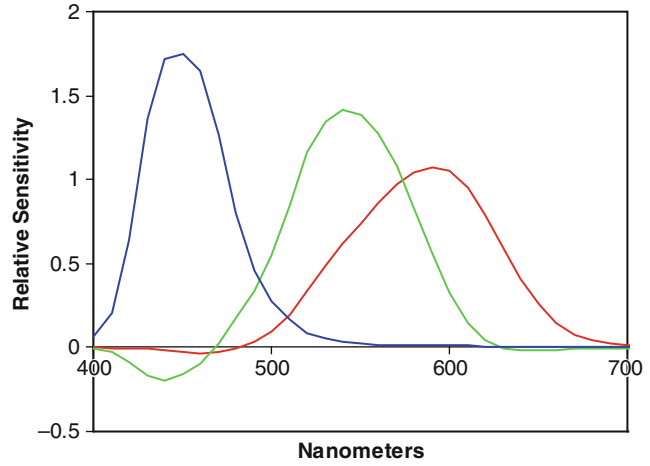
The result of applying this matrix to 1931 color matching functions is shown in Fig. 3. These curves can be compared to the CIE color matching functions, and it can be seen that these curves are qualitatively more narrow or sharpened. This matrix was derived using a set of corresponding color training data but excluding highly chromatic light sources.

The degree of adaptation is the next step in the model and is calculated given the surround setting



CIECAM02, Fig. 2 Overall flowchart of the computational steps for calculating perceptual attribute correlates given input tristimulus values and a specific set of viewing conditions

CIECAM02, Fig. 3 The CAT02 matrix shown as a set of *red*, *green*, and *blue* sensitivity curves computed by applying Eq. 1 to the CIE 1931 color matching functions



and the luminance of the adapting field. The specific calculation for incomplete adaptation or D is shown in Eq. 2. The variable L_A is the luminance of the adapting field and the value of F is a parameter that is computed from the surround setting.

$$D = F \left[1 - \left(\frac{1}{3.6} \right) e^{\left(\frac{-L_A - 42}{92} \right)} \right] \quad (2)$$

The results of using Eq. 2 with three different surround settings over luminance of adapting field values ranging from 0 to 800 cd/m^2 are shown in Fig. 4. Essentially the surround limits the degree of adaptation and increases with larger values of L_A . Complete adaptation is only achieved with the average surround with high L_A values.

Given the input stimulus converted to CAT02 RGB space and a degree of adaptation, it is then possible to compute the chromatic adaptation. This can be done according to Eqs. 3, 4, and 5 below. The stimulus CAT02 values are shown as R , G , and B and the adopted or adapted white point as R_w , G_w , and B_w .

$$R_c = [D(100/R_w) + 1 - D]R \quad (3)$$

$$G_c = [D(100/G_w) + 1 - D]G \quad (4)$$

$$B_c = [D(100/B_w) + 1 - D]B \quad (5)$$

The next step in the calculation of the forward CIECAM02 model is shown as the box HPE RGB

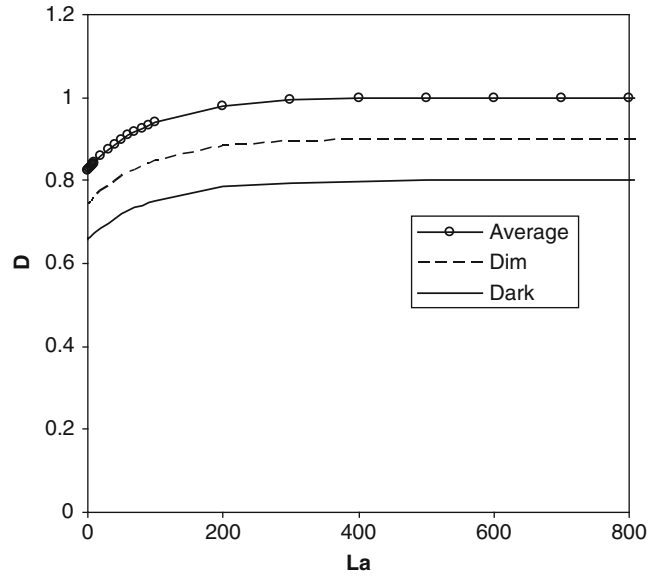
in the center of the flowchart in Fig. 2. This is the conversion of R_c , G_c , and B_c above to the Hunt-Pointer-Estevéz space. This can be done using a 3 by 3 matrix shown in Eq. 6:

$$\mathbf{M}_{HPE} = \begin{bmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix} \quad (6)$$

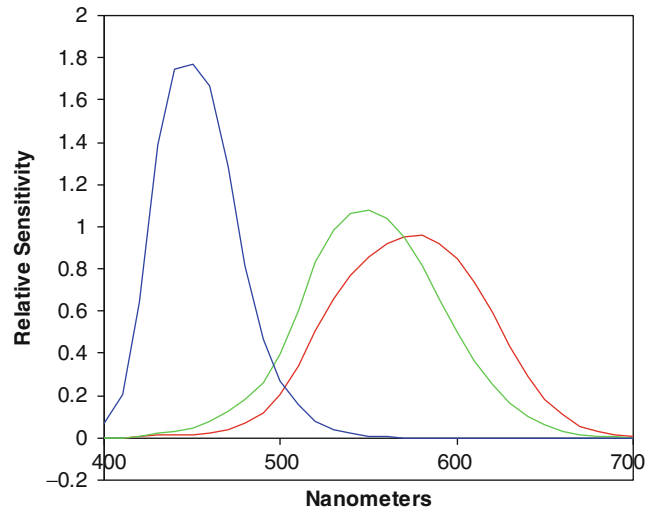
A plot of the corresponding red, green, and blue sensitivity curves for the Hunt-Pointer-Estevéz RGB space is shown in Fig. 5. This graph shows the red and green curves with qualitatively broader and more overlapping than the red and green curves for the CAT02 RGB curves shown in Fig. 3. Effort was made during the formulation of the model to derive a single RGB space for both the chromatic adaptation and for the nonlinear compression, but the results were generally worse and ultimately the final model made use of the two different RGB spaces.

Given the Hunt-Pointer-Estevéz RGB values, the next step is the nonlinear response compression. The specific nonlinearity used [6] in CIECAM02 is shown in Eqs. 7 through 9 and is shown graphically in Fig. 6. The R' , G' , and B' values are the HPE values as computed with Eq. 6, and the F_L value is a model parameter that is dependent on the viewing conditions.

CIECAM02, Fig. 4 The degree of adaptation as computed for three surrounds and varying L_A . A D value of 1 is complete adaptation to the white point, while values less than one are for incomplete adaptation



CIECAM02, Fig. 5 The HPE matrix shown as a set of *red*, *green*, and *blue* sensitivity curves computed by applying Eq. 6 to the CIE 1931 color matching functions



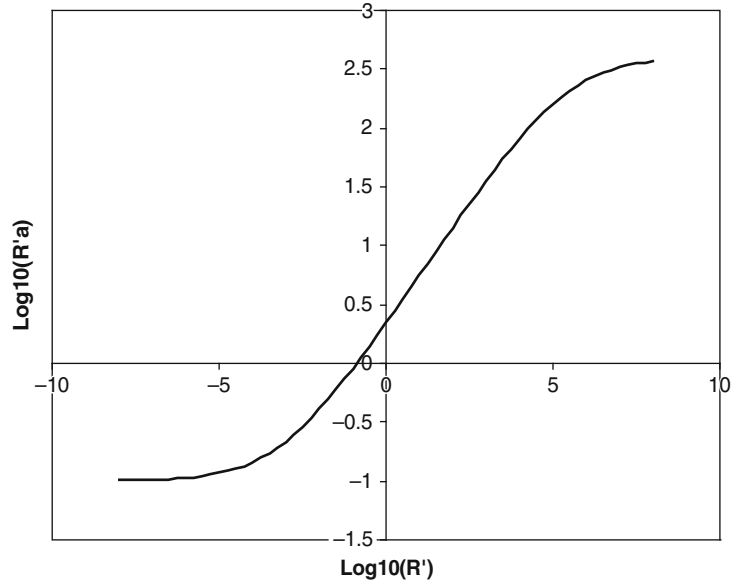
$$R'_a = \frac{400(F_L R' / 100)^{0.42}}{[27.13 + (F_L R' / 100)^{0.42}]} + 0.1 \quad (7)$$

$$G'_a = \frac{400(F_L G' / 100)^{0.42}}{[27.13 + (F_L G' / 100)^{0.42}]} + 0.1 \quad (8)$$

$$B'_a = \frac{400(F_L B' / 100)^{0.42}}{[27.13 + (F_L B' / 100)^{0.42}]} + 0.1 \quad (9)$$

The final steps in the forward CIECAM02 model are the computation of the perceptual attribute correlates. First an intermediate set of opponent coordinates, a and b , are computed according to Eqs. 10 and 11. It should be emphasized that these values of a and b are preliminary and should not be used directly. The value of h or hue is computed using the arctangent of b divided by a . A table of constants is used to compute H or hue quadrature. The resulting H values for red, yellow, green, and blue are 100, 200, 300, and 400, respectively.

CIECAM02, Fig. 6 The post-adaptation nonlinear response compression function as computed using Eq. 7. Similar curves result for Eqs. 8 and 9 for the calculation of G'_a and B'_a



$$a = R'_a - 12G'_a/11 + B'_a/11 \quad (10)$$

$$b = (1/9)(R'_a + G'_a - 2B'_a) \quad (11)$$

The perceptual attribute correlates for lightness and brightness can then be calculated. First the achromatic signal or A is computed according to Eq. 12. The R'_a , G'_a , and B'_a values are the non-linearly compressed values from Eqs. 7 through 9. Next the computation of J or lightness is shown in Eq. 13, while the computation of Q or brightness is shown in Eq. 14. The c and z values are additional model parameters as computed based on the viewing conditions.

$$A = [2R'_a + G'_a + (1/20)B'_a - 0.305]N_{bb} \quad (12)$$

$$J = 100(A/A_w)^{cz} \quad (13)$$

$$Q = (4/c)\sqrt{J/100}(A_w + 4)F_L^{0.25} \quad (14)$$

Finally given the correlates for hue, lightness, and brightness, it is possible to calculate the perceptual attribute correlates for chroma, colorfulness, and saturation. First a temporary variable t is computed using Eq. 15. Next C or chroma is calculated using Eq. 16. Colorfulness or M and saturation s can then be computed using Eqs. 17 and 18.

$$t = \frac{(50,000/13)N_c N_{cb} e_t (a^2 + b^2)^{1/2}}{R'_a + G'_a + (21/20)B'_a} \quad (15)$$

$$C = t^{0.9} \sqrt{J/100} (1.64 - 0.29^n)^{0.73} \quad (16)$$

$$M = CF_L^{0.25} \quad (17)$$

$$S = 100\sqrt{M/Q} \quad (18)$$

The result of the preceding calculations is then a set of perceptual attribute correlates for the given input values and viewing conditions. Note though that this does not define a rectangular coordinate system, such as the a^* and b^* values for CIELAB [7]. Instead a set of correlates such as lightness, chroma, and hue must first be computed and used as polar coordinates. The rectangular coordinates can be computed using Eqs. 19 and 20. Similar coordinates can be calculated for lightness and saturation, with subscript s , and brightness and colorfulness, with subscript M .

$$a_c = C \cdot \cos(h) \quad (19)$$

$$b_c = C \cdot \sin(h) \quad (20)$$

The inverse CIECAM02 equations are beyond the scope of this entry, but reference 1 contains the full steps for inverting the above equations. This

allows tristimulus values to be computed given input perceptual attribute correlates and viewing conditions.

The CIECAM02 color appearance model can then be used in situations requiring a viewing condition independent color encoding. It also could be considered in cases where CIELAB or CIELUV [7] lacks perceptual uniformity. For example, the CIELAB blue hue lacks hue constancy and tends toward purple as it approaches the neutral axis. This can be problematic in many circumstances, such as when gamut mapping colors from a display to a printer. The CIECAM02 space is considerably more uniform in this case. Two example gradients are shown in Fig. 7. The CIELAB gradient shown on the top has a clear tendency to purple as it goes to gray while the CIECAM02 gradient on the bottom does not. Both color spaces have constant hue angles for these colors but CIECAM02 is significantly improved.

It is useful to further compare and contrast CIELAB and CIECAM02. CIELAB has as input the stimulus XYZ values and the white point XYZ values. CIECAM02 has as input the stimulus and white XYZ values and also luminance of the adapting field, the luminance of the background, and the surround setting. CIELAB has a chromatic adaptation transform that consists of a complete von Kries transform in XYZ space, while CIECAM02 has a complete or incomplete von Kries transform in CAT02 RGB space. CIELAB has a cube-root nonlinearity, while CIECAM02 uses a modified hyperbolic function as the nonlinearity. CIELAB uses XYZ data to compute the opponent signal, while CIECAM02 is based on a

Hunt-Pointer-Estevéz space. Finally CIELAB can be used to compute lightness, chroma, and hue correlates, while CIECAM02 can be used to compute these values as well as brightness, colorfulness, saturation, and hue quadrature values. However, CIELAB has benefited from the additional research in advanced color difference equations and as a result has advanced color difference metrics such as ΔE_{94} and ΔE_{2000} which CIECAM02 does not have. There are encouraging results [8] though for using CIECAM02-based color difference equations.

Future Directions

CIECAM02 has been a useful and valuable addition to color appearance modeling research. It has provided a single reference point for ongoing research in the area of color appearance modeling. However, a number of researchers have pointed to specific aspects of the complexity that are problematic in some cases. For example, for the darkest colors, it may not be possible to invert the calculations for highly saturated inputs. These values may be outside the spectral locus, but for color management applications that use a fixed intermediate grid of coordinates, this is a shortcoming. Therefore, it seems likely that future work will continue in the area of color appearance modeling, with a future focus on robustness [8] and perhaps simplicity. In spite of these limitations, there is already work integrating CIECAM02 with color management systems, such as the International Color Consortium (ICC) [10, 11]. There has also been work [12] to



CIECAM02, Fig. 7 *Light gray to deep blue constant hue angle gradients for CIELAB, shown at the top, and CIECAM02 shown at the bottom. The left- and right-hand*

side colors are the same XYZ values and have constant respective hue angles, but the top gradient for CIELAB tends to a purplish tone in the center of the gradient

consider how the model could be further extended to encompass a wider range of viewing conditions, such as mesopic illumination levels. Finally, there is also research [13] in the area of how the model could be used with complex stimuli to create an image appearance model.

Cross-References

- [CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02](#)
- [CIELAB](#)

References

1. CIE Publication 159: A Colour Appearance Model for Colour Management Systems: CIECAM02. (2004)
2. Fairchild, M.D.: A revision of CIECAM97s for practical applications. *Color. Res. App.* **26**, 418–427 (2001)
3. Hunt, R.W.G., Li, C.J., Juan, L.Y., Luo, M.R.: Further improvements to CIECAM97s. *Color. Res. App.* **27**, 164–170 (2002)
4. Fairchild, M.D.: *Color Appearance Models*. Addison Wesley, Reading (1998)
5. Hunt, R.W.G.: Models of color vision. In: *Measuring Color*, pp. 208–247. Fountain Pr. Ltd, Kingston-upon-Thames (2001)
6. Hunt, R.W.G., Li, C.J., Luo, M.R.: Dynamic cone response functions for models of colour appearance. *Color. Res. App.* **28**, 82–88 (2003)
7. CIE Pub. No. 15: *Colorimetry*. Central Bureau of the CIE, Vienna (2004)
8. Li, C.J., Luo, M.R., Cui, G.: Colour-differences evaluation using colour appearance models. In: *Proceedings of the IS&T/SID Color Imaging Conference*, pp. 127–131, Scottsdale, Arizona, USA (2003)
9. Li, C.J., Luo, M.R.: Testing the robustness of CIECAM02. *Color. Res. App.* **30**, 99–106 (2005)
10. Kuo, C., Zeise, E., Lai, D.: Robust CIECAM02 implementation and numerical experiment within an international color consortium workflow. *J. Imaging Sci. Technol.* **52**, 020603–020606 (2008)
11. Tastl, I., Bhachech, M., Moroney, N., Holm, J.: ICC color management and CIECAM02. In: *Proceedings of the IS&T/SID Color Imaging Conference*, pp. 217–223, Scottsdale, Arizona, USA (2005)
12. Kwak, Y., MacDonald, L.W., Luo, M.R.: Mesopic colour appearance. *Proc. SPIE*. **5007**, 161–169 (2003)
13. Tulet, O., Larabi, O., Fernandez-Maloigne, M.C.: Image rendering based on a spatial extension of the CIECAM02. In: *IEEE workshop application of computer vision*, Anchorage, Alaska, USA (2008)

CIECAM02 (Standards: CIE)

- [Gamut Volume](#)

CIEDE2000, History, Use, and Performance

Ming Ronnier Luo

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Synonyms

[CIE 2000 color-difference equation](#); ΔE_{00}

Definition

CIEDE2000 [1, 2] is a color-difference formula recommended by the CIE in year 2001. It has also recently been published as an ISO and CIE Joint Standard [3]. It is a modification of CIELAB [4] and gives an overall best performance in predicting experimental datasets. The typical applications are pass/fail decision, color constancy, metamerism, and color rendering.

Overview

Over the years, color scientists and engineers have been striving to achieve a single number pass/fail color-difference equation, i.e., to apply a single pass/fail color difference to all color regions for industrial quality control. In practice, product batches should be visually acceptable against a standard, when a color difference is less than a predetermined color-difference unit (called color tolerance). Reversely, it will be rejected to be returned for re-shading.

In 1976, CIELAB uniform color space was recommended by the CIE. The decision was

made based on limited experimental data. It was realized a shortage of reliable experimental data having similar color-difference magnitude as those used in industry (typically with CIELAB color difference (ΔE^*_{ab}) ≤ 5). Hence, many datasets were produced, in which four datasets were considered to be most comprehensive and robust and were used to derive the new formulae. These datasets are Luo and Rigg [5], RIT-DuPont [6], Witt [7], and Leeds [8]. They have 2776, 156, 418, and 307 pairs of samples and average

color differences of 3.0, 1.0, 1.9, and 1.6 ΔE^*_{ab} , respectively. All datasets were based on glossy paint samples except that of Luo and Rigg data, which include many subsets based on different materials, and finally all subsets were combined according to the experimental results based on textile samples [5].

Using these data, a series of equations were developed by modifying CIELAB. They all have a generic form as given in Eq. 1:

$$\Delta E = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C_{ab}^*}{k_C S_C} \frac{\Delta H_{ab}^*}{k_H S_H}\right)} \quad (1)$$

where

$$\begin{aligned} \Delta L^* &= L_{ab,2}^* - L_{ab,1}^* \\ \Delta C_{ab}^* &= C_{ab,2}^* - C_{ab,1}^* \\ \Delta H_{ab}^* &= 2\sqrt{C_{ab,2}^* C_{ab,1}^*} \sin\left(\frac{\Delta h_{ab}}{2}\right) \\ \text{where } \Delta h_{ab} &= h_{ab,2} - h_{ab,1} \end{aligned}$$

and subscripts 1 and 2 are the two samples in a pair. The R_T is an interactive term between ΔC_{ab}^* and ΔH_{ab}^* . The S_L , S_C , and S_H are weighting factors for the correlates of L^* , C_{ab}^* , and h_{ab} and are dependent on the positions of the samples in a pair. The k_L , k_C , and k_H are parametric factors to take into account the surface characteristics of the materials in question such as textile, paint, and plastic. Equation 1 is a form of ellipsoid along the directions of CIELAB lightness, chroma, and hue angle. The ellipsoid can also be rotated in C_{ab}^* and h_{ab} plane. Four equations were developed after CIELAB in 1976. These were Leeds [8], BFD [9], CIE94 [10] and CMC [11]. The CMC was adopted by the ISO for textile applications in 1992 [11]. The CIE94 was recommended by the CIE for field trials in 1994. Both equations have the first three terms of Eq. 1, and BFD and Leeds include all four terms. All formulae greatly outperform CIELAB to fit the experimental datasets. Industry was confused which equation

should be used. Hence, CIE Technical Committee (TC) 1–47 *Hue and Lightness-Dependent Correction to Industrial Colour Difference Evaluation* was formed in 1998. It was hoped that a generalized and reliable formula could be achieved.

The TC members worked closely together and a formula named CIEDE2000 was recommended [1, 2]. The computation procedure of this formula is given in Eq. 2.

CIEDE2000 ($K_L:K_C:K_H$) Color-Difference Formula

The input of the equation is two sets of CIELAB values for the samples of the pair in question. The procedure to calculate CIEDE2000 is given below:

Step 1. Prepare data to calculate a' , C' , and h' .

$$L' = L^*$$

$$a' = (1 + G)a^*$$

$$b' = b^*$$

$$C'_{ab} = \sqrt{a'^2 + b'^2}$$

$$h_{ab} = \tan^{-1} \left(\frac{b'}{a'} \right)$$

where

$$G = 0.5 \left(1 - \sqrt{\frac{\overline{C_{ab}^*}^7}{\overline{C_{ab}^*}^7 + 25^7}} \right)$$

where $\overline{C_{ab}^*}$ is the arithmetic mean of the C_{ab}^* values for a pair of samples.

Step 2. Calculate $\Delta L'$, $\Delta C'$, and $\Delta H'$.

$$\Delta L' = L'_2 - L'_1$$

$$\Delta C'_{ab} = C'_{ab,2} - C'_{ab,1}$$

$$\Delta H'_{ab} = 2 \sqrt{C'_{ab,2} C'_{ab,1}} \sin \left(\frac{\Delta h'_{ab}}{2} \right)$$

$$\text{where } \Delta h'_{ab} = h'_{ab,2} - h'_{ab,1}$$

Step 3. Calculate CIEDE2000 ΔE_{00} .

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L} \right)^2 + \left(\frac{\Delta C'_{ab}}{k_C S_C} \right)^2 + \left(\frac{\Delta H'_{ab}}{k_H S_H} \right)^2 + R_T \left(\frac{\Delta C'_{ab}}{k_C S_C} \right) \left(\frac{\Delta H'_{ab}}{k_H S_H} \right)}$$

where

$$S_L = 1 + \frac{0.015 (\overline{L'} - 50)^2}{\sqrt{20 + (\overline{L'} - 50)^2}}$$

and

$$S_C = 1 + 0.045 \overline{C'_{ab}}$$

and

$$S_H = 1 + 0.015 \overline{C'_{ab}} T$$

where

$$T = 1 - 0.17 \cos(\overline{h'_{ab}} - 30^\circ) + 0.24 \cos(2\overline{h'_{ab}}) + 0.32 \cos(3\overline{h'_{ab}} + 6^\circ) - 0.20 \cos(4\overline{h'_{ab}} - 63^\circ)$$

and

$$R_T = -\sin(2\Delta\theta) R_C$$

where

$$\Delta\theta = 30 \exp \left\{ - \left[\left(\overline{h'_{ab}} - 275^\circ \right) / 25 \right]^2 \right\}$$

$$\text{and } R_C = 2 \sqrt{\frac{\overline{C'_{ab}}^7}{\overline{C'_{ab}}^7 + 25^7}}$$

(2)

Note that $\overline{L'}$, $\overline{C'_{ab}}$, and $\overline{h'_{ab}}$ are the arithmetic means of the L' , C'_{ab} , and h'_{ab} values for a pair of samples. For calculating the $\overline{h'_{ab}}$ value, caution needs to be taken for neutral colors having hue angles in different quadrants, e.g., Sample 1 and

Sample 2 with hue angles of 90° and 300° would have a mean value of 195° , which differs from the correct answer, 15° . This can be obtained by checking the absolute difference between two hue angles. If the difference is less than 180° , the arithmetic mean should be used. Otherwise, 360°

should be subtracted from the larger angle, followed by calculating of the arithmetic mean. This gives $300-360^\circ = -60^\circ$ for the sample and a mean of $(90-60^\circ)/2 = 15^\circ$ in this example.

Three-Term CIEDE2000 Color-Difference Formula

The CIEDE2000 color-difference equation described above includes four terms. In many applications, a three-term equation is required such as for shade sorting and color tolerance specification, indicating direction to a specific difference in recipe prediction. Hence, a three-term CIEDE2000 version was also developed [12] and is Eq. 3:

$$\Delta E_{00} = \left[(\Delta L_{00})^2 + (\Delta C_{00})^2 + (\Delta H_{00})^2 \right]^{1/2} \quad (3)$$

where

$$\Delta L_{00} = \frac{\Delta L'}{k_L S_L}; \quad \Delta C_{00} = \frac{\Delta C''}{S_C''}; \quad \Delta H_{00} = \frac{\Delta H''}{S_H''}$$

and

$$\begin{aligned} \Delta C'' &= \Delta C' \cos(\varphi) + \Delta H' \sin(\varphi) \\ \Delta H'' &= \Delta H' \cos(\varphi) - \Delta C' \sin(\varphi) \end{aligned}$$

$$\text{where } \tan(2\varphi) = R_T \frac{(k_C S_C) (k_H S_H)}{(k_H S_H)^2 - (k_C S_C)^2}$$

where φ is taken to be between -90° and 90° . If $k_H S_H = k_C S_C$, then 2φ is equal to 90° and φ is equal to 45° :

$$\begin{aligned} S_C'' &= (k_C S_C) \sqrt{\frac{2(k_H S_H)}{2(k_H S_H) + R_T(k_C S_C) \tan(\varphi)}} \\ S_H'' &= (k_H S_H) \sqrt{\frac{2(k_C S_C)}{2(k_C S_C) - R_T(k_H S_H) \tan(\varphi)}} \end{aligned}$$

where $\Delta L'$, $\Delta C'$, and $\Delta H'$ and S_L , S_C , S_H , k_L , k_C , and k_H are the same as those in Eq. 2.

The ΔE_{00} value calculated from Eq. 3 is the same as calculated from Eq. 2.

Evaluation of the CIEDE2000

The typical way to evaluate each formula's performance is to apply statistical measures. One measure, standardized residual sum of squares (STRESS) index [13], has been widely used as given in Eq. 4. It measures the discrepancy between the calculated color difference (ΔE) and visual difference (ΔV) from an experimental dataset:

$$\text{STRESS} = 100 \sqrt{\frac{\sum (\Delta E_i - f \Delta V_i)^2}{f^2 \sum \Delta V_i^2}} \quad (4)$$

$$\text{where } f = \frac{\sum \Delta E_i^2}{\sum \Delta E_i \Delta V_i}$$

STRESS values are ranged between 0 and 100. For a perfect agreement, STRESS value should be zero. It can be considered as a percentage error prediction. Table 1 gives the performance of the four equations tested using the previous mentioned four datasets together with COM dataset which was combined with a weight for each of the four datasets.

It can be seen in Table 1 that CIEDE2000 gave an overall best performance. In addition, CIELAB performed the worst. CIEDE2000 performed significantly better than the other formulae except insignificantly better than CIE94 for the Leeds and RIT-DuPont sets.

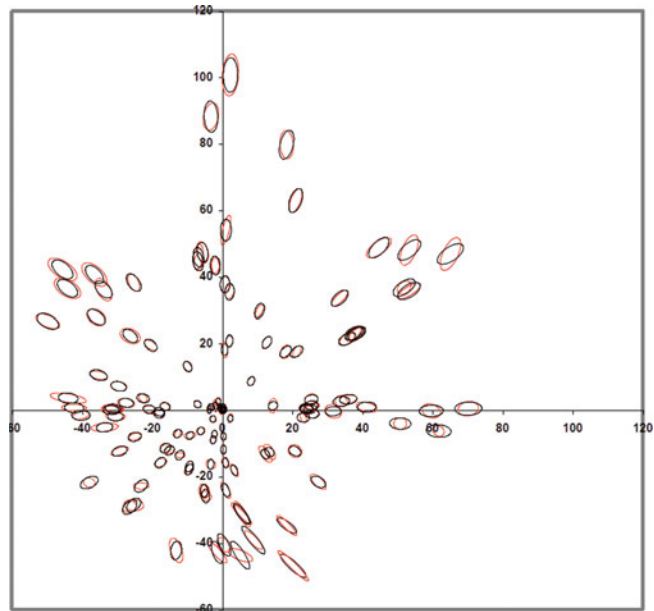
Another widely used method to evaluate color-difference equations is to use color discrimination ellipses. Figure 1 shows that the experimental ellipses (in black color) obtained from the Luo and Rigg dataset plotted against those (in red color) predicted by CIEDE2000. The color discrimination ellipse is an effective way to represent a number of color-difference pairs in a color center. The color differences between points in the ellipse and the color center represent equal visual color difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. Hence, the pattern shown in Fig. 1 indicates a poor performance of CIELAB, i.e., very small ellipses close to neutral axis; ellipse sizes increase when chroma increases.

CIEDE2000, History, Use, and Performance, Table 1 Color-difference formula performance in STRESS unit (Copyright of the Society of Dyers and Colourists)

	COM	BFD	Leeds	RIT-DuPont	Witt
CIELAB	44	42	40	33	52
CIE94	32	34	21	20	32
CIEDE2000	27	30	19	19	30

CIEDE2000, History, Use, and Performance,

Fig. 1 Experimental color discrimination ellipses plotted in a^*b^* diagram. (Copyright of the Society of Dyers and Colourists)



Comparing the experimental and CIEDE2000 ellipses, both sets fit well, especially in the blue region. (Note that the experimental ellipses for all other color regions generally point toward the neutral except that the blue ellipses point away from the neutral axis. This effect implies that both CMC and CIE94 formulae do fit well to the experimental results in the region.)

Future Directions

This section showed an outstanding color-difference equation, CIEDE2000, which has been recommended by the CIE. It fits the datasets having magnitude of industrial color differences well. However, it does not have an associated color space. A summary of the future direction on color difference is given below.

- Almost all of the recent efforts have been spent on the modifications of CIELAB. This has resulted in CIEDE2000 including five corrections of CIELAB to fit the available experimental datasets. It is desirable to derive a formula based upon a new perceptually uniform color space from a particular color vision theory such as CIECAM02 [14].
- All color-difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color-difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size of samples, size of color difference, separation, and background. Again, CIECAM02 model and its extension CAM02-UCS [15] are derived to follow this direction.
- Almost all of the color-difference formulae were developed only to predict the color

difference between a pair of large single objects/patches. More and more applications require to predict color differences between a pair of pictorial images. The current formula does not include necessary components to consider spatial variations for evaluating images. Johnson and Fairchild developed a spatial model based on CIEDE2000 [16].

Cross-References

- [CIE94, History, Use, and Performance](#)
- [CIECAM02](#)
- [CIELAB](#)

References

1. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. *Color. Res. Appl.* **26**, 340–350 (2001)
2. Pub, C.I.E., No.: 142.: Improvement to Industrial Colour-Difference Evaluation. Central Bureau of the CIE, Vienna (2001)
3. ISO 11664–6:2008(E)/CIE S 014–6/E.: Joint ISO/CIE Standard: Colorimetry-Part 6: CIEDE2000 (2007)
4. CIE Publ. 15: Colorimetry. Central Bureau of the CIE, Vienna (2004)
5. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. *Color. Res. Appl.* **11**, 25–42 (1986)
6. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of supra-threshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
7. Witt, K.: Geometric relations between scales of small colour differences. *Color. Res. Appl.* **24**, 78–92 (1999)
8. Kim, H., Nobbs J.H.: New weighting functions for the weighted CIELAB colour difference formula, *Proc. Colour 97 Kyoto*. 1, 446–449 (1997)
9. Luo, M.R., Rigg, B.: *BFD(l:c)* colour difference formula, part I—development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
10. CIE: Industrial Colour-Difference Evaluation, CIE Publ.116. Central Bureau of the CIE, Vienna (1995)
11. ISO 105-J03: Textiles: Test for Colour Fastness. Part 3 Calculation of Colour Differences. ISO, Geneva (2009)
12. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008)
13. Nobbs, J.H.: A lightness, chroma and hue splitting approach to CIEDE2000 colour differences. *Adv. Colour. Sci. Technol.* **5**, 46–53 (2002)
14. Pub, C.I.E., No.: 159.: A Colour Appearance Model for Colour Management Systems: CIECAM02. Centre Bureau of the CIE, Vienna (2004)
15. Luo, M.R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. *Color. Res. Appl.* **31**(320–330), 425–435 (2006)
16. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. *Color. Res. Appl.* **28**, 425–435 (2003)

CIELAB

Ming Ronnier Luo

State Key Laboratory of Modern Optical

Instrumentation, Zhejiang University, Hangzhou, China

School of Design, University of Leeds, Leeds, UK

Synonyms

[CIE 1976 L*a*b*](#); [DE*ab](#); [History, use and performance](#)

Definition

CIELAB is a uniform color space (UCS) recommended by CIE in 1976 [1], and it was later published as a Joint ISO/CIE Standard [2]. A UCS is defined by the CIE International Lighting Vocabulary [3] as a color space in which equal distances are intended to represent threshold or suprathreshold perceived color differences of equal size. It is one of the most widely used color spaces. The typical applications include color specification and color difference evaluation. The former is to describe a color in perceptual correlates such as lightness, chroma, and hue and to plot samples to understand their relationships. The latter is mainly used for color quality control such as setting color tolerance, color constancy, metamerism, and color rendering.

For the definition equations of the components of CIELAB, see section CIE L*a*b* Formula (CIELAB).

Overview

Over the years, color scientists and engineers have been striving to achieve a UCS. To apply UCS, a pair of samples will first be measured by a color measuring instrument to obtain their CIE tristimulus values (XYZ) which will then be transformed to the perceptual correlates such as CIELAB lightness, chroma, and hue angle. The distance between a pair of colors is calculated and reported as color difference (ΔE). This difference will then be judged against a predetermined color tolerance which could be a particular color region and a product. For a particular product, all pairs should be judged as “acceptable,” when the color difference is less than the color tolerance. Otherwise, it will be rejected. A good color difference formula is also called a “single number pass/fail formula,” i.e., to apply a single color tolerance to all color regions.

Over 20 formulae were derived before the recommendation of CIELAB in 1976 [4]. Some of them were derived to fit the spacing of the Munsell color order system. The concept of the Munsell color order system was invented by A. H. Munsell and was based on steps of equal visual perception. Any color can be defined as a point in

a three-dimensional Munsell color space. Its associated attributes are Munsell hue (H), Munsell chroma (C), and Munsell value (V) which correspond to the perceived hue, saturation, and lightness, respectively. The spacing of the color samples for each attribute was intensively studied by the members in the Colorimetry Committee of the Optical Society of America (OSA), and the CIE tristimulus values of ideally spaced samples were published in 1943 [5].

Figure 1 shows the loci for samples having constant Munsell chroma and curves for samples having constant Munsell hue. Since the Munsell samples are based on equal visual steps, for a perfect agreement between the Munsell data and a color space, all loci should be circles with a constant increment between all neighboring chroma steps. As shown in Fig. 1, this is obviously not the case, i.e., one step of chroma in the blue region is at least five times shorter than one step of chroma in the green region. Additionally, all iso-chroma loci are far from being circles, and no iso-hue contours are straight lines. These indicate that there is a very large difference between the Munsell system and the CIE system represented by x, y chromaticity diagram.

Some earlier Munsell-based formulae are directly calculated using Munsell H, V, and C values with a weighting factor for each component. In 1944, ANLAB was developed by Adams and Nickerson [6] as given in Eq. 1.

$$\Delta E_{\text{ANLAB}} = \sqrt{40 \left\{ (0.23 \Delta V_y)^2 + [\Delta(V_x - V_y)]^2 - 0.4 [\Delta(V_y - V_z)]^2 \right\}} \quad (1)$$

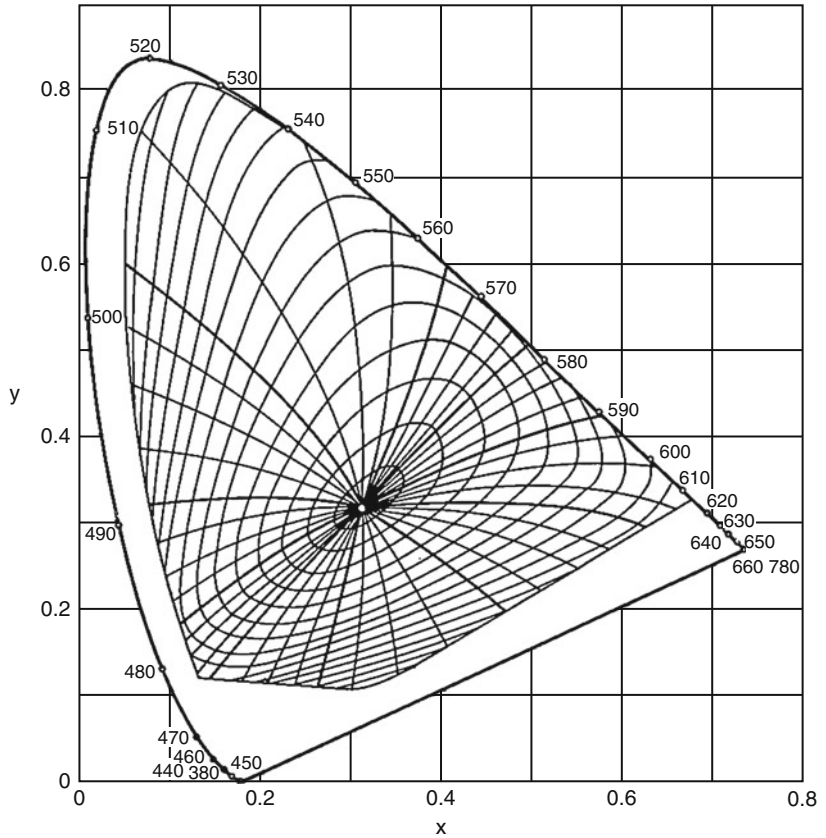
and

$$I = 1.2219 V_1 - 0.23111 V_1^2 + 0.23591 V_1^3 \\ - 0.021009 V_1^4 + 0.00084045 V_1^5$$

where I corresponds to X , Y , or Z tristimulus values. In Eq. 1, the terms of $(V_x - V_y)$ and $0.4(V_y - V_z)$ correspond to the ANLAB

a (redness-greenness) and b (yellowness-blueness) scales, respectively. By adding the third scale $0.23V_y$, ANLAB becomes a three-dimensional UCS. It was recommended by the Colour Measurement Committee (CMC) of the Society of Dyers and Colourists (SDC) and became an ISO standard in 1971 for the application in the textile industry. A series of cube root formulae were also derived to simplify the

CIELAB, Fig. 1 Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIE chromaticity diagram. (From Billmeyer and Saltzman [14])



ANLAB formula which involves a cumbersome fifth-order polynomial function. This resulted in CIELAB color difference formula introduced in 1976 [1]. CIELAB units include L^* , a^* , and b^* ; the asterisk is used to differentiate the CIELAB system from ANLAB.

In 1976, the CIE recommended two uniform color spaces, CIELAB (or CIE $L^*a^*b^*$) and CIELUV (or CIE $L^*u^*v^*$), as it was still not possible to decide which one would correspond better to visual observations.

CIE $L^*a^*b^*$ Formula (CIELAB)

CIELAB equation is given in Eq. 2.

$$\begin{aligned} L^* &= 116f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \quad (2)$$

where

$$f(I) = I^{1/3}, \text{ for } I > \left(\frac{6}{29}\right)^3$$

Otherwise,

$$F(I) = \frac{841}{108}I + \frac{16}{116}$$

X, Y, Z and X_n, Y_n, Z_n are the tristimulus values of the sample and a specific reference white considered. It is common to use the tristimulus values of a CIE standard illuminant as the X_n, Y_n, Z_n values. Correlates of L^* , a^* , and b^* are lightness, redness-greenness, and yellowness and blueness, respectively.

Correlates of hue and chroma are also defined by converting the rectangular a^*, b^* axes into polar coordinates (see Eq. 3). The lightness (L^*), chroma (C_{ab}^*), and hue (h_{ab}) correlates

correspond to perceived color attributes, which are generally much easier to understand for describing colors.

$$\begin{aligned} h_{ab} &= \tan^{-1}(b^*/a^*) \\ C^*_{ab} &= \sqrt{a^{*2} + b^{*2}} \end{aligned} \quad (3)$$

Color difference can be calculated using Eq. 4.

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (4)$$

or

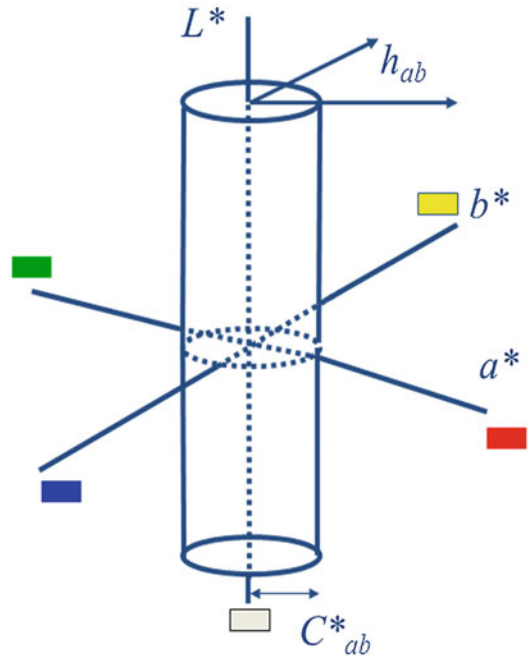
$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta C^{*2}_{ab} + \Delta H^{*2}_{ab}}$$

where

$$\Delta H^*_{ab} = 2 \left(C^*_{ab,1} C^*_{ab,2} \right)^{1/2} \sin \left[\frac{(h_{ab,2} - h_{ab,1})}{2} \right] \quad (5)$$

and subscripts 1 and 2 represent the samples of the pair considered; the ΔL^* , Δa^* , Δb^* , ΔC^*_{ab} , and ΔH^*_{ab} are the difference of L^* , a^* , b^* , C^*_{ab} , and hue in radiant unit (see Eq. 5) between Samples 1 and 2, respectively.

Figure 2 shows the CIELAB color space. It can be seen that the rectangular coordinates consist of L^* , a^* , and b^* . A positive and negative values of a^* represent reddish and greenish colors, respectively. A positive and negative values of b^* represent yellowish and bluish colors, respectively. For the polar coordinates, hue angle is ranged from 0° to 360° following a rainbow scale from red, yellow, green, blue, and back to red. The 0° , 90° , 180° , and 270° approximate pure red, yellow, green, and blue colors (or unitary hues). Chroma starts from zero origin of neutral axis having chroma of zero and then increases its chromatic content to become more colorful. The colors located in the cylinder of Fig. 4 have a constant chroma, equally perceived chroma content around the hue circle.



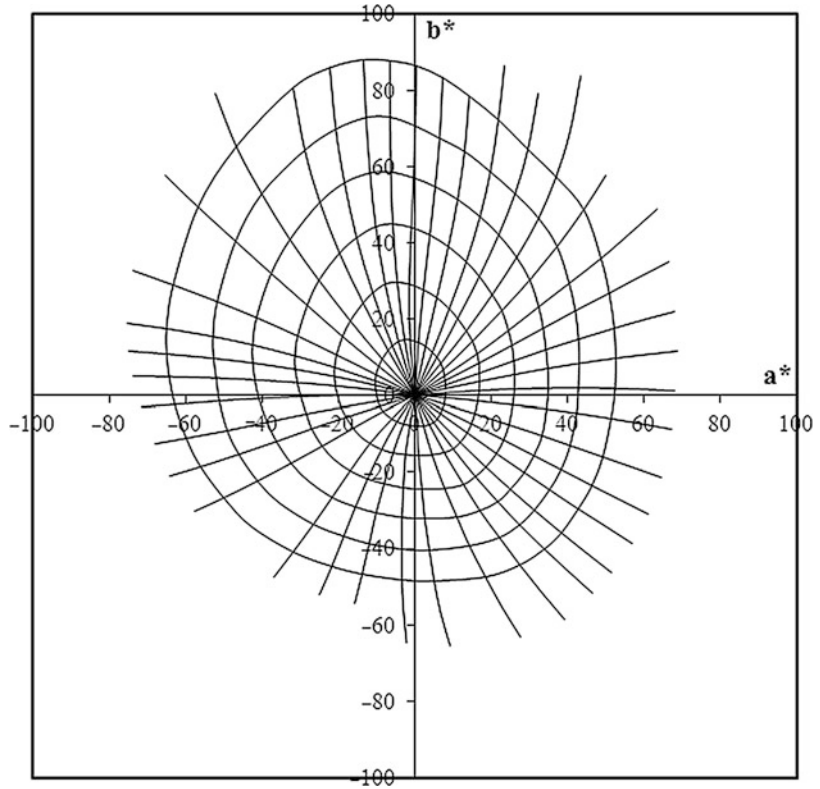
CIELAB, Fig. 2 Illustration of CIELAB color space

Evaluation of CIELAB Color Space

Two sets of data are used here to evaluate the performance of CIELAB space. Figure 3 shows the constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in CIELAB a^*b^* diagram. It can be seen that the pattern is close to the expectation of a good UCS, i.e., the constant chroma loci are close to circle and constant hue radius are close to a straight line (hue constancy). The uniformity of CIELAB is much better than that of CIE chromaticity diagram (see Fig. 1). However, detailed inspection can be found that the same chroma value in yellow and blue regions could still differ by a factor of almost 2. Also, the constant hue line is very much curved in the areas of orange, blue, and green yellow.

After the recommendation of the CIELAB formula, some experimental datasets having $\Delta E^*_{ab} \leq 5$ (representing typical magnitude of industrial color differences) were produced. These sets in general agree with each other. They were later selected to be used to derive different

CIELAB, Fig. 3 Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIELAB a^*b^* diagram



color difference formulae. Figure 4 shows the experimental ellipses obtained from the Luo and Rigg dataset [7]. The dataset includes more color centers and covers a much larger color gamut than the others. A color discrimination ellipse represents the points on the circumference of the ellipse against the center of the ellipse to have the same visual difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. The figure shows the poor performance of CIELAB. For example, all ellipses close to neutral are much smaller than those in the saturated color regions.

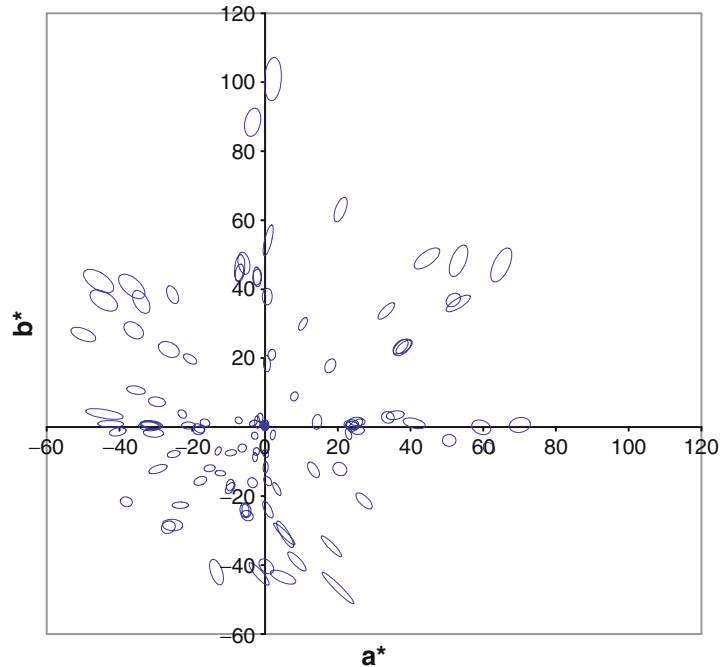
The results given in Figs. 3 and 4 clearly showed the effect of performance that different experimental results could disagree with each other greatly. The discrepancy between the Luo and Rigg and Munsell datasets is mainly due to the color difference magnitude used in the experimental datasets. CIELAB performs not badly for the large color differences with ΔE^*_{ab} about 10 units (see Fig. 3), but predicts very poorly for smaller color differences ($\Delta E^*_{ab} \leq 5$) (see Fig. 4).

Future Directions

Since the recommendation of CIELAB in 1976, many equations were derived by modifying CIELAB such as CMC [8], CIE94 [9], and the more recent CIE recommendation CIEDE2000 [10]. They do fit the datasets ($\Delta E^*_{ab} \leq 5$) well. However, they do not have an associated color space. The future directions on color difference are given below.

- It is desirable to derive a formula based upon a new perceptually uniform color space from a model of color vision theory such as CIECAM02 [11]. A uniform color space is based upon this color appearance model, like CAM02-UCS [12].
- All color difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size

CIELAB, Fig. 4 Luo and Rigg experimental color discrimination ellipses plotted in $a^* b^*$ diagram



of samples, size of color difference, separation, and background. Again, the CIECAM02 model [11] and its extension, CAM02-UCS [12], are equipped with these capabilities.

- Almost all of the color difference formulae were developed only to predict the color difference between a pair of individual patches. More and more applications require evaluating color differences between a pair of pictorial images. Johnson and Fairchild developed a formula for this purpose [13].

Cross-References

- CIE 2000 Color-Difference Equation
- CIE Tristimulus Values
- CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space
- CIE94, History, Use, and Performance
- CIECAM02
- CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

References

1. CIE Publication No. 015: Colorimetry. Central Bureau of the CIE, Vienna (2004)
2. ISO 11664-4:2008(E)/CIE S 014-4/E: Joint ISO/CIE Standard: Colorimetry-Part4: CIE 1976 $L^*a^*b^*$ Colour Space (2007)
3. CIE: International Lighting vocabulary, Central Bureau of the CIE, Vienna. <http://eilv.cie.co.at/> (2012)
4. Luo, M. R.: The development of colour-difference formulae. *Rev. Prog. Color. SDC*. 28–39 (2002)
5. Newhall, S.M., Nickerson, D., Judd, D.B.: Final report of the O.S.A. subcommittee on spacing of the Munsell colors. *J. Opt. Soc. Am.* **33**, 385–418 (1943)
6. Adams, E.Q.: X-Z planes in the 1931 ICI system of colorimetry. *J. Opt. Soc. Am.* **32**, 168–173 (1942)
7. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. *Color. Res. Appl.* **11**, 25–42 (1986)
8. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dye. Colour.* **100**, 128–132 (1984). and 281–282
9. CIE Publication No. 116: Industrial colour-difference evaluation. Central Bureau of the CIE, Vienna (1995)
10. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. *Color. Res. Appl.* **26**, 340–350 (2001)
11. CIE Publication No. 159: A colour appearance model for colour management systems: CIECAM02. (2004)

12. Luo, M.R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. *Color. Res. Appl.* **31**, 320–330 (2006)
13. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. *Color. Res. Appl.* **28**, 425–435 (2003)
14. Billmeyer, F.W., Saltzman, M.: *Principles of Colour Technology*, 2nd edn. Wiley, New York (1981)

CIELAB (Standards: CIE)

► Gamut Volume

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

Robert R. Buckley¹ and Edward J. Giorgianni²

¹National Archives of the UAE, Abu Dhabi, UAE

²Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

Synonyms

CIE $L^*a^*b^*$; CIELab; ICC $L^*a^*b^*$; ICCLAB; ITULAB; Lab

Definition

Using the L^* , a^* , and b^* coordinates of the CIELAB uniform color space as the three component values of a 24-bit digital image.

CIELAB

Overview

CIE 1976 (L^* , a^* , b^*) color space (or CIELAB) is widely used for image encoding and processing applications as diverse as color management, gamut mapping, color interchange, and color quality evaluation. What makes CIELAB

attractive for these applications is its being an approximately uniform color space. A perceptually uniform scale generally enables the most efficient use of bits and the least visible artifacts when images are digitized for viewing by human observers.

CIELAB Color Space

In CIELAB color space a color stimulus is expressed in terms of L^* , a^* , and b^* rectangular coordinates. The entry on CIELAB gives the formulas for calculating L^* , a^* , b^* values from the X , Y , Z tristimulus values of the color stimulus and the X_n , Y_n , Z_n tristimulus values of the reference white object color stimulus. CIELAB color space can be traced to the Munsell system, which was based on evaluating suprathreshold color differences of isolated reflective patches viewed against a gray background.

In the CIELAB color space, the L^* value represents the lightness value of the stimulus. A stimulus that reflects no radiation over the visible wavelength band has an L^* value of zero, while the perfectly diffusing reflector under the illuminant with tristimulus values X_n , Y_n , Z_n has an L^* value of 100. Stimuli with tristimulus values that are proportional to those of the reference white color stimulus will have a^* and b^* values of zero. This means that CIELAB is an opponent color space: a gray or achromatic stimulus with the same chromaticities as the normalizing white stimulus will lie on the L^* axis, with black at one end and white at the other. The cube root formula for L^* is a close approximation of the relationship between the Munsell value V and luminance factor Y/Y_n , so L^* is essentially the Munsell value scaled up by a factor of 10. The distance of a stimulus from the L^* axis is given by the chroma value C^*_{ab} .

In an ideal perceptually uniform color space, the Euclidean distance between two color stimuli would correspond to the perceived difference between them. Ideally, then, pairs of color stimuli would look equally different if the three-dimensional distance between them in CIELAB color space, i.e., ΔE^*_{ab} , was the same, independent of the difference being along the L^* , a^* , or b^* dimension or some combination of them. A ΔE^*_{ab}

of 1 is generally considered a just perceptible difference, although the threshold depends on how the stimuli are compared and what part of color space they lie in because CIELAB is only approximately uniform. Since 1976, the CIE has published two extensions to the CIE 1976 $L^*a^*b^*$ color-difference formulas to improve perceptual uniformity in the calculation of color differences, i.e., CIE94 and CIEDE2000.

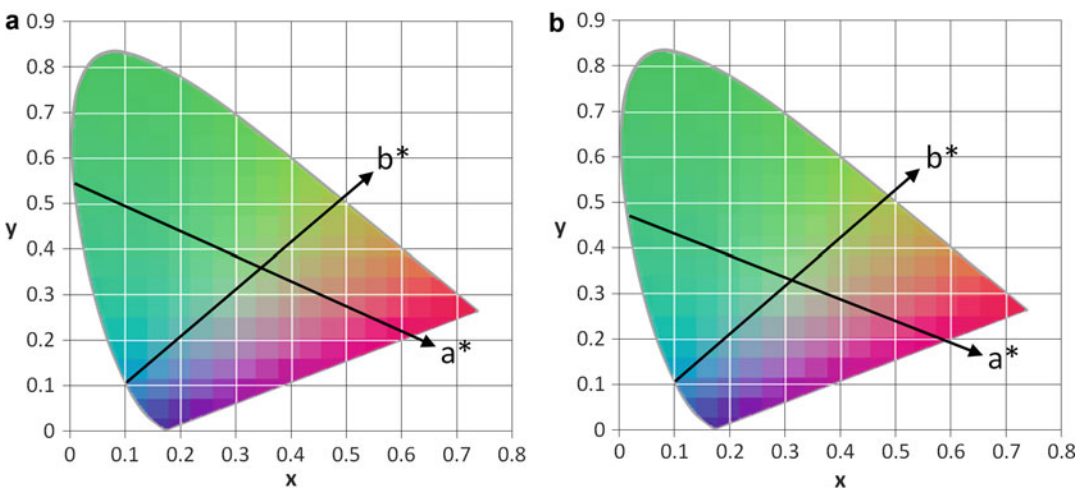
Figure 1 plots the a^* and b^* axes for CIE standard illuminants D_{50} and D_{65} on the x-y chromaticity diagram. In very rough terms, the a^* axis runs from cyan to magenta and the b^* axis, from blue to yellow. The range of a^* and b^* values is determined by the collection of samples at hand or that a given device can generate. Fig. 2 shows selected L^* planes of the volume generated in CIELAB color space by the MacAdam limits for reflectance materials with respect to CIE standard illuminants D_{50} and D_{65} . For the L^* planes shown in Fig. 2, the a^* values range from -163 to 126.6 for illuminant D_{50} and from -168.2 to 145.6 for illuminant D_{65} ; the b^* values range from -129.6 to 145.1 for illuminant D_{50} and from -127.7 to 145.3 for illuminant D_{65} . These are theoretical limits on the values that can be obtained with reflectance materials.

The ranges obtained with object color stimuli encountered in practice with printed and

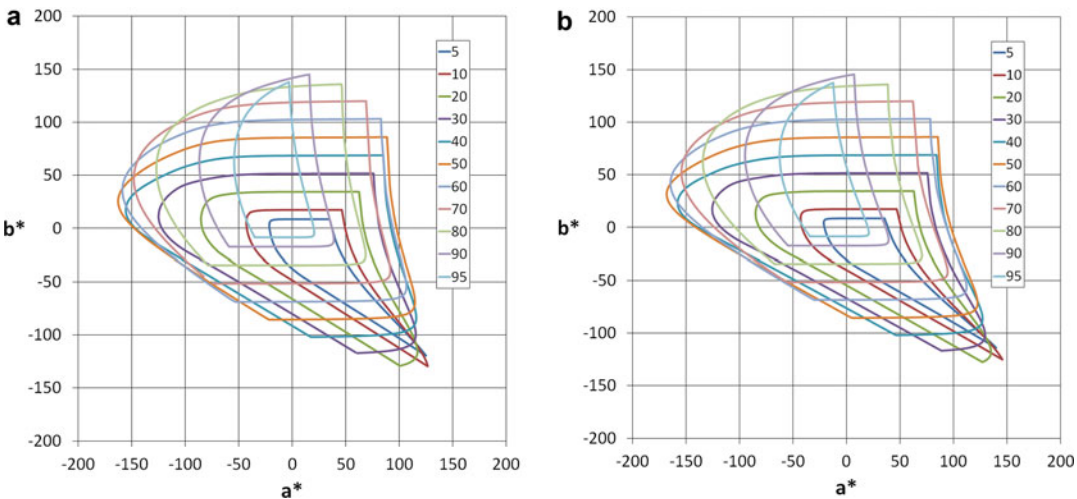
photographic materials will be much less than the MacAdam limits. For example, Fig. 3 shows the Pointer colors measured with illuminant C, which has a correlated color temperature close to that of CIE illuminant D_{65} . An update to the Pointer gamut with a revised set of stimuli and for CIE illuminant D_{50} is under development as this is being written.

Quantization and Dynamic Range

A desirable property of a color space for image encoding is that it covers as large a gamut as needed with the fewest number of levels or digital values. Too many levels are wasteful; too few and an image will show visible contours or bands due to the lower precision in the quantization. Banding is most noticeable in areas with smooth gradients. If the sampling of values or quantization in color space is not fine enough, small differences in the original color values of adjacent pixels in such areas will become larger and potentially visible color differences after quantization. The usual solution is to ensure the quantization is fine enough, especially in the regions of color space where the human visual system is most sensitive to color differences. The addition of noise also can help. If the quantization is mathematically uniform, then quantizing the color space everywhere the same means that the quantization



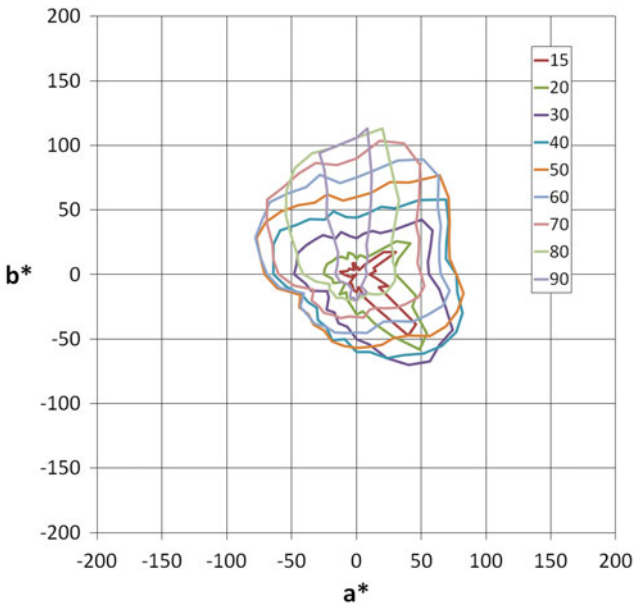
CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 1 a^* , b^* axes on a CIE x-y chromaticity diagram with (a) CIE standard illuminant D_{50} and (b) CIE standard illuminant D_{65} as normalizing illuminant



CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 2 a^* and b^* values of the MacAdam limits for L^* planes from 5 to 95 for (a) CIE standard illuminant D_{50} and (b) CIE standard illuminant

D_{65} . (Figure is based on data generously provided by Prof. Francisco Martinez-Verdú of the Colour & Vision Group at the University of Alicante)

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 3 a^* and b^* values of the Pointer colors for L^* planes from 15 to 90



will be finer than it needs to be in regions of color space less sensitive to color differences, which is inefficient.

The optimum solution for a uniform color quantization of rendered images is to use a perceptually uniform color space. (*Rendered images* are defined here to mean images intended to be directly viewed by human observers.) This

solution would use the fewest levels for a given limit on the maximum color error in the quantization or sampling process. This performance is achieved with a three-dimensional quantization based on a body-centered cubic lattice [1, 2], which would also give the minimum mean squared error if the input color stimuli were uniformly distributed [3]. A one-dimensional

quantization based on a simple cubic lattice and that samples each CIELAB coordinate L^* , a^* , and b^* independently would be suboptimal but more straightforward to implement. Another reason for less than optimal performance is that CIELAB is only approximately uniform. The divergence of CIELAB from true perceptual uniformity has led to extensions to improve upon various nonuniformities. Nevertheless, the original CIELAB color space is still widely used for many color imaging applications.

For one-dimensional quantization, 8 bits are sufficient to quantize the range of a^* or b^* values for reflection images typically encountered in practice. While 7 bits would be enough to represent numerical values from 0 to 100, the use of 8 bits generally would be more convenient in a digital system. Moreover, there are perceptual grounds for using a minimum of 8 bits to quantize the L^* coordinate in color images. As mentioned earlier, isolated color patches were used in the development of the Munsell system to which CIELAB color space can be traced. When such patches abut along a sharp edge, smaller color differences are detectable. There are historical data [4] and more recent analysis [5] to justify scaling the L^* value up to 8 bits so that the just noticeable differences along the L^* , a^* , and b^* axes across a sharp edge are more nearly equal. Also, the greater weighting given to L^* differences in color-difference equations in some applications recognizes a generally greater sensitivity to luminance differences compared to chromatic differences.

While there may be room for discussion and debate about the exact value of the scaling factor to use for the lightness coordinate, from the beginning standard engineering practice has been to use 8 bits when encoding L^* in color images [2, 6]. Some CIELAB applications, especially those in which CIELAB-encoded images may later be subjected to significant amounts of additional image processing, may instead specify 12 or 16 bits per coordinate.

CIELAB image data is generated by converting linear tristimulus or image values. Since CIELAB is a nonlinear transform of

(linear) XYZ tristimulus values, 12 bits of tristimulus values are needed to address all 256 levels in an 8-bit CIELAB system in order to minimize quantization errors and banding. Nonlinear 8-bit-per-coordinate RGB encodings such as sRGB can be converted to CIELAB using a combination of matrices and one-dimensional lookup tables as long as sufficiently greater bit precision is used in the intermediate conversion stages. Three-dimensional lookup tables are often used for such conversions.

Because CIELAB was developed to be a mathematical representation of the Munsell system – an approximately uniform space based on reflective color patches – it follows that CIELAB should be well suited for encoding images on reflection media having luminance dynamic ranges and chromaticity boundaries similar to those of the Munsell system color set. In practice, this often is the case. Color images on most reflection media can be encoded satisfactorily in terms of 24-bit (8 bits per channel) CIELAB values if those values are used judiciously. In particular, the luminance dynamic range of the image medium generally should not exceed a visual density range of about 2.50 (a luminance ratio of just over 300:1). The exact luminance dynamic range that can be encoded in terms of 8-bit CIELAB L^* values will depend on a number of factors, most notably the amount of noise in the images. The greater the noise, the less visible quantization artifacts such as contouring will be. In many imaging systems, the noise inherent in the original images and the additional noise contributed by scanners, display devices, and media are sufficient to mask quantization artifacts. However, in other cases, such as virtually noise-free digitally created images, noise in the form of dithering may need to be computationally added in order to avoid unacceptable quantization when 8-bit-per-channel CIELAB encoding is used.

Encoding in terms of 8-bit-per-channel CIELAB values sometimes can be problematic even when applied to images on reflection media or other imaging media or devices having comparable luminance dynamic ranges. In particular, undesirable banding and other artifacts may result

if the encoded images are subjected to significant additional image processing. For example, if an image encoded at 8-bit-per-channel CIELAB is appreciably lightened or darkened by subsequent signal processing, quantization effects that originally may have been subthreshold may become visible in image shadow or highlight areas of the processed image. In an extreme case, image processing might be used to reverse the sign of an image, e.g., to make a positive original image into a negative image. In that case, the L^* quantization of the original image would no longer be appropriate for the transformed image, and undesirable quantization effects likely would result.

Another consideration for 8-bit-per-channel CIELAB encoding, as well as for virtually any other 8-bit-per-channel encoding, is its applicability for rendered images having luminance dynamic ranges or chromaticity gamuts significantly greater than those of the Munsell system. For example, imaging devices such as digital projectors and media such as photographic slide films and motion picture films have luminance dynamic ranges considerably greater than those of typical reflection media. When the perceived lightness of such images is first appropriately adjusted for observer general brightness adaptation and then expressed in terms of CIELAB values, the resulting L^* values can range from well below 1.0 to well over 100 [7]. The luminance dynamic ranges of high dynamic range (HDR) digital projectors and other HDR display devices are even greater, and the primaries of such devices can be capable of producing chromaticity color gamuts much larger than that of the Munsell system. In addition, images to be encoded may be colorimetric representations of original scenes rather than of rendered reproductions. Unrendered scene colorimetric values may be in the form of camera RAW values from a digital camera, computationally derived values from scans of color negatives or other photographic media, or values produced by computer-generated imaging (CGI). Such values can greatly exceed those that can be adequately represented by virtually any 24-bit encoding metric, including 8-bit-per-channel color encodings based on CIELAB and other color spaces derived from the Munsell system.

CIELAB Encodings

Different image formats and the applications that use them have made different encoding choices for CIELAB. One of the first standards to define a CIELAB encoding for image and document processing applications was the Xerox Color Encoding Standard [8]. Since then the TIFF, PDF, and JPX formats also have defined CIELAB encodings.

An obvious way to digitally encode L^* , a^* , and b^* values is to convert them to integer values. The TIFF CIELAB encoding [9] does this, using an unsigned integer value for L^* and signed integer values for a^* and b^* . Other encodings have been designed that use unsigned integers and map the L^* , a^* , and b^* values encountered in practice to match the 8-bit range so as to reduce quantization errors by making best use of the 8 bits available for encoding the values.

When CIELAB is used for image encoding, then the two main choices are:

- What does the L^* range represent?
- What mapping should be used to encode the a^* and b^* values?

The following equations map L^* , a^* , and b^* values to 8-bit encoded L , A , and B values:

$$\begin{aligned} L &= \text{scale}_L \times L^* + \text{offset}_L \\ &= \frac{255}{(L_{\max}^* - L_{\min}^*)} (L^* - L_{\min}^*) \end{aligned}$$

$$\begin{aligned} A &= \text{scale}_A \times a^* + \text{offset}_A \\ &= \frac{255}{(a_{\max}^* - a_{\min}^*)} (a^* - a_{\min}^*) \end{aligned}$$

$$\begin{aligned} B &= \text{scale}_B \times b^* + \text{offset}_B \\ &= \frac{255}{(b_{\max}^* - b_{\min}^*)} (b^* - b_{\min}^*) \end{aligned}$$

The encodings can be described using either scale and offset parameters or the minimum and maximum values that are mapped to the 8-bit values 0 and 255. In some formats, the encoding parameters are defined by the format; in others,

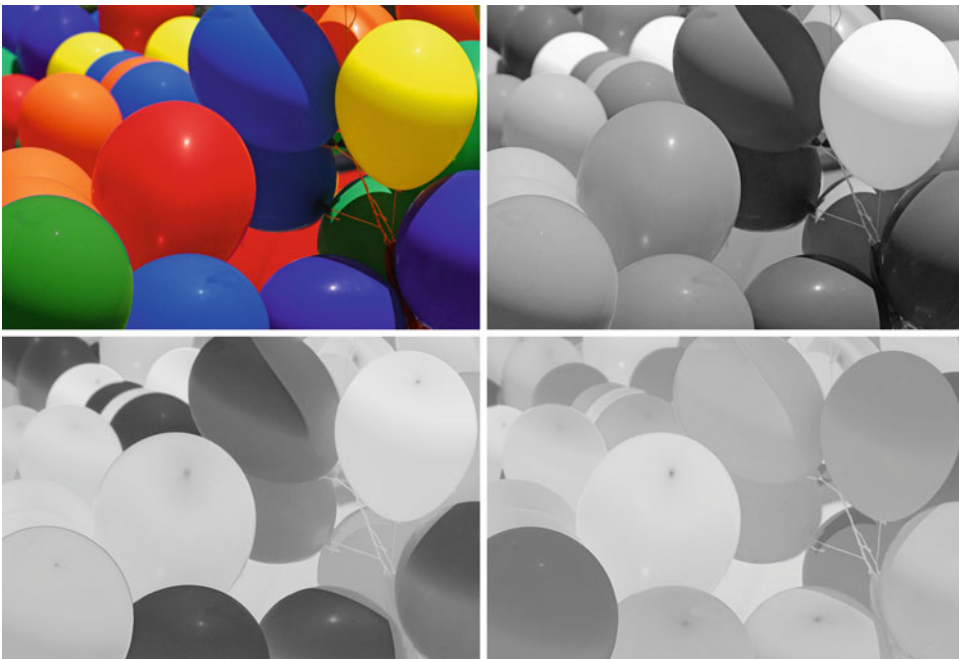
the user can explicitly set them with default values usually available.

In most formats the L scale factor $scale_L$ is 255/100 and the L offset parameter $offset_L$ is 0. The equivalent minimum and maximum L^* values are 0 and 100. This means that L^* values of 0 and 100 are encoded as 0 and 255. This is the default L^* encoding in TIFF and PDF. In these formats, an L^* value of 100 corresponds to the perfect diffusing reflector under the specified illuminant. In the case of the TIFF CIELAB encoding, the illuminant is not part of the encoding specification. PDF requires that the diffuse white point be specified. Specifying the illuminant is optional in JPX, which uses CIE illuminant D_{50} as the default. A question sometimes asked is what illuminant should be used with CIELAB. The answer is the reference illuminant used to measure or calibrate the XYZ or RGB values from which the CIELAB values were derived.

The a^* and b^* values typically encountered in practice lie within the 8-bit range from -128 to 127 so they can be encoded using a scale factor of

1 and an offset of 0, which is what the TIFF CIELAB encoding does [9]. The encoded A and B values that result are signed 8-bit integers. The TIFF ICCLAB encoding [10] uses an offset value of 128 instead so that the encoded A and B values are unsigned 8-bit integers, as they are in the 8-bit encoding that the International Color Consortium (ICC) profile format standard uses when CIELAB is the profile connection space [11]. Figure 4 shows the encoded L , A , and B components for the TIFF ICCLAB encoding of a sample image, where a^* and b^* values of 0 are encoded as 128.

Since the ranges of typical a^* and b^* values are usually much less than 255, a scale factor greater than 1 (or equivalently a maximum-minimum difference of less than 255) would expand the range encountered in practice to fill the full range available with 8 bits. The TIFF ITULAB encoding [12] allows an application to specify the minimum and maximum values of a^* and b^* that map to the encoded values 0 and 255. This encoding matches the CIELAB encoding used by color facsimile [13]. PDF [14] provides the same CIELAB encoding for images. JPX [15], the extended



CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 4 Clockwise from *upper left*, original color image and its L^* , a^* , and b^* components encoded using the TIFF ICCLAB encoding

version of the JP2 base file format for JPEG 2000, provides an equivalent capability by allowing an application to specify the ranges of a^* and b^* values that map onto the range of 8-bit values and the offsets to go with them.

Besides using the full range available with an 8-bit encoding, scaling the a^* and b^* values also reduces quantization errors, especially when image compression is used. This was important for color facsimile, where reasonable transmission times require the use of moderate to high levels of JPEG compression, which can lead to visible “blockiness” in the decompressed image. Scaling the a^* and b^* values reduces the effects of the quantization error caused by aggressive JPEG compression. The default scaling defined in the ITU color facsimile standards was based on the range obtained by measuring a wide range of hardcopy materials; the range was -85 to $+85$ for a^* and -75 to $+125$ for b^* . The default illuminant for the CIELAB encoding is CIE illuminant D_{50} . The color facsimile standards allow color fax machines to negotiate the use of custom a^* and b^* ranges and other illuminants.

The case for scaling up the L^* values encountered in practice to fill the full 8-bit range is less compelling since the values encountered in practice already cover most of the available range. For example, the minimum and maximum reflectance values for the reference medium defined by the ICC are 0.30911% and 89%, corresponding to a visual density range of about 2.46 (a luminance ratio of about 288:1) and minimum and maximum L^* values of 2.79 and 95.58, using the perfectly diffusing reflector as the white reference. While higher dynamic ranges with L^* values greater than 100 originating from scenes and photographic

films can be modified to fit within an 8-bit range, the results can be loss of detail due to the tone compression and limited precision provided by 8 bits in those cases.

Table 1 compares 8-bit CIELAB encodings in terms of the defaults for minimum and maximum L^* , a^* , and b^* values and whether the format allows an application to set the ranges for L^* , a^* , or b^* . While the focus here is on 8-bit encodings, these formats also support 16-bit CIELAB encodings.

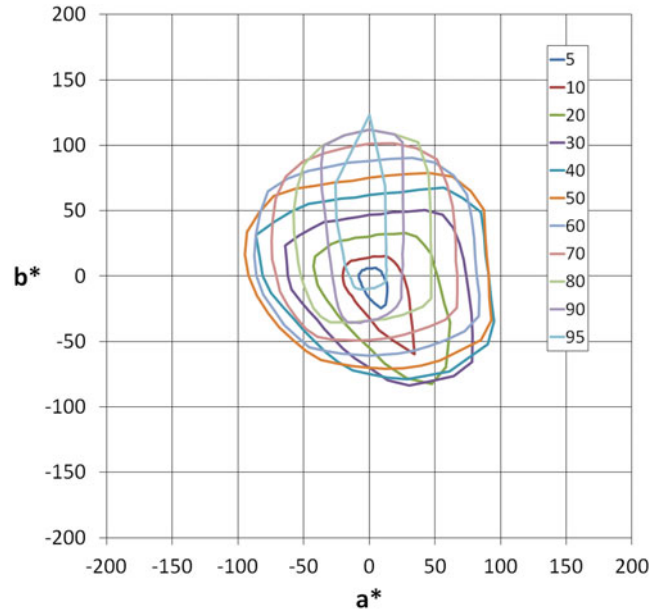
In the CIELAB encodings described thus far, an L^* value of 100 corresponds to the perfectly diffusing reflector under the specified illuminant. This is not always the case with the CIELAB encoding that the ICC defines for the profile connection space (PCS). The PCS is the color space to or from which input and output device coordinates can be transformed. (Besides CIELAB, the PCS may also use an XYZ encoding.) This makes it possible to connect an arbitrary input device to an arbitrary output device via the PCS in an ICC-based color management system, since a device only needs to know how to transform its coordinates to or from PCS coordinates. Such transforms are defined in device profiles. Transforms can have different intents such as perceptual, where the objective is a pleasing reproduction, or colorimetric, where the objective is a colorimetrically accurate reproduction of the tristimulus values.

Lookup table-based profiles for the *perceptual* rendering intent on input, output, and display devices all use CIELAB as the PCS; both 8- and 16-bit versions are defined. The CIELAB values of the PCS are calculated with XYZ tristimulus values that are normalized with respect to the media white point under CIE illuminant D_{50} rather

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Table 1 Default component ranges for 8-bit CIELAB encodings

Format – encoding	L^*		a^*		b^*		Range settable
	Min	Max	Min	Max	Min	Max	
TIFF – CIELAB	0	100	-128	127	-128	127	No
TIFF – ICCLAB	0	100	-128	127	-128	127	No
TIFF – ITULAB	0	100	-85	85	-75	125	L^*, a^*, b^*
PDF – Lab	0	100	-100	100	-100	100	a^*, b^*
JPX – CIELab	0	100	-85	85	-75	125	L^*, a^*, b^*

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 5 a^* and b^* values of the ISO 12640-3 reference color gamut for L^* planes from 5 to 95



than the perfectly diffusing reflector. (The PCS CIELAB values for the *colorimetric* intent are calculated with reference to the perfectly diffusing reflector.) In keeping with the orientation to graphic arts and desktop publishing applications, V4 of the ICC specification [11] defined the reference medium for calculating the PCS values of the perceptual rendering as a reflection print on a substrate with a neutral reflectance of 89% and where the darkest printable color has a neutral reflectance of 0.30911%. Therefore, the L^* value of the unprinted substrate is 100 and of the darkest printable color is 3.1373. With the 8-bit CIELAB encoding used for the PCS, these values are scaled and encoded as 255 and 8, respectively. A scale factor of 1 and an offset of 128 are used for the a^* and b^* values, as in the TIFF ICCLAB encoding.

The adoption of a reference medium, and subsequently of a reference medium gamut, was designed to improve interoperability within an ICC-based workflow using the perceptual intent. Because a perceptual transform from source to destination media usually required gamut mapping, the gamut of the reference medium used for the PCS was explicitly defined. Besides specifying the endpoints of the L^* scale of the perceptual reference medium, this meant specifying the a^* and b^* limits as well. For this the ICC used the

reference color gamut defined in ISO 12640-3 [16] and shown in Fig. 5. As with the PCS, the L^* , a^* , and b^* values in Fig. 5 were calculated using the reference medium (instead of the perfectly diffusing reflector) with CIE illuminant D_{50} as the reference white object color in the L^* , a^* , and b^* equations. The ISO 12640-3 standard also specifies a set of TIFF (specifically TIFF/IT) images with 16-bit CIELAB values that use the ICC PCS encoding and are within the limits of the reference gamut in Fig. 5.

Conclusions

Based on this discussion, it can be concluded that 8-bit-per-channel CIELAB is appropriate and well suited for digitally encoding color images when all of the conditions below apply:

- The images to be encoded are rendered reproductions.
- The luminance dynamic range and overall color gamut of the images are consistent with those of the Munsell system.
- Appropriate mathematical transformations, such as scale factors and offsets, are used to

maximize the efficiency of the CIELAB encoding.

- Relatively limited subsequent image processing is applied to CIELAB-encoded images.

Cross-References

- [CIE Tristimulus Values](#)
- [CIE94, History, Use, and Performance](#)
- [CIEDE2000, History, Use, and Performance](#)

References

1. Buckley, R.: The quantization of CIE uniform color spaces using cubic lattices. In: *Colour 93; 7th Congress of the AIC*. pp. 246–247. International Colour Association (AIC), Budapest (1993)
2. Buckley, R.: Color Image coding and the geometry of color space. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge (1981)
3. Barnes, E.S., Sloane, N.J.A.: The optimal lattice quantizer in three dimensions. *SIAM J. Algebraic Discrete Methods*. **4**, 30–41 (1983)
4. Nickerson, D., Newhall, S.M.: A psychological color solid. *J. Opt. Soc. Am.* **33**, 419–422 (1943)
5. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
6. Schreiber, W.F., Buckley, R.R.: A two-channel picture coding system: II – adaptive companding and color coding. *IEEE Trans. Commun.* **COM-29**, 1849–1858 (1981)
7. Giorgianni, E.J., Madden, T.E.: *Digital Color Management: Encoding Solutions*, 2nd edn, pp. 235–237. Wiley, Chichester (1998)
8. Xerox Systems Institute: *Xerox Color Encoding Standard XN55 288811*, Sunnyvale (1989)
9. Adobe Systems Inc: *TIFF, Revision 6.0*, Mountain View (1992)
10. Adobe Systems Inc: *Adobe Pagemaker® 6.0 TIFF Technical Notes*, Mountain View (1995)
11. International Color Consortium: *Specification ICC.1: 2010 (Profile Version 4.3.0.0) Image technology colour management – Architecture profile format and data structure* http://color.org/specification/ICC1v43_2010-12.pdf (2010). Accessed 2 Jan 2013.
12. Internet Engineering Task Force (IETF): *RFC 3949, File Format for Internet Fax* (2005)
13. International Telecommunication Union (ITU): *ITU-T Recommendation T.42 Continuous tone colour representation method for facsimile* (2003)
14. International Organization for Standardization (ISO): *ISO 32000-1:2008, Document management – Portable document format – Part 1: PDF 1.7* (2008)
15. International Organization for Standardization (ISO): *ISO 15444-2: 2004, Information technology – JPEG 2000 image coding system: Extensions* (2004)
16. International Organization for Standardization (ISO): *ISO 12640-3:2007, Graphic technology – Prepress digital data exchange – Part 3: CIELAB standard colour image data (CIELAB/SCID)*

Circadian Rhythms

- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)

Circadian System

- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)

Clapper-Yule Model

Mathieu Hébert

ERIS Group, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet de Saint-Etienne, Saint-Etienne, France

Definition

The Clapper-Yule model is a physically based model describing the reflection of spectral light fluxes by a printed surface and enabling the prediction of halftone prints on diffusing substrates [1]. The model relies on a closed-form equation obtained by describing the multiple transfers of light between the substrate and the print-air interface through the inks. Physical parameters are attached to the inks, the diffusing support, and the surface. The model assumes that the lateral propagation distance of light within the substrate, due to scattering, is much larger than the halftone screen period. Most photons therefore cross different ink dots while traveling in the print. The

reflections and transmissions of light at the surface are explicitly taken into account depending on the print's refractive index, as well as the considered illumination and measuring geometries.

The Clapper-Yule Equation

The Clapper-Yule equation derives from the description of multiple reflections of light between the substrate and the print-air interface, represented in Fig. 1. The substrate, strongly diffusing, has a spectral reflectance $r_g(\lambda)$. The print-air interface reflects on its two sides: On the top side, a fraction r_s of the incident light is reflected toward the detector (this fraction may be zero according to the geometry of illumination and detection). A fraction T_{in} enters the print. On the back side, a fraction r_i is reflected and a fraction T_{ex} exits the print toward the detector. Regarding the inks, the model assumes that they are located between the surface and the substrate. If N different inks are used, the halftone is a mosaic of 2^N colors, also called *Neugebauer primaries*, resulting from the partial overlap of the ink dots. In the case of typical cyan, magenta, and yellow inks, there are eight Neugebauer primaries: white (no ink), cyan, magenta, yellow, red (magenta + yellow), green (cyan + yellow), blue (cyan + magenta), and black

(cyan + magenta + yellow). Each primary occupies a fractional area a_i on the surface and transmits a fraction $t_i(\lambda)$ of the light. The global attenuation for light crossing the halftone ink layer is therefore:

$$\sum a_i t_i(\lambda).$$

By denoting, $I_0(\lambda)$ and $I_1(\lambda)$ the downward fluxes illuminating the surface, respectively, the substrate, and as $J_0(\lambda)$ and $J_1(\lambda)$ the upward fluxes exiting the surface, respectively, the substrate, the following relations are obtained:

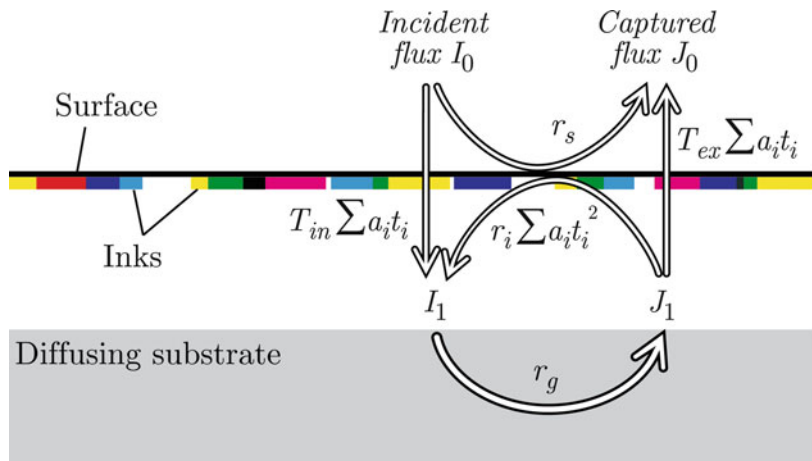
$$\begin{aligned} J_0(\lambda) &= r_s I_0(\lambda) + \left[T_{ex} \sum a_i t_i(\lambda) \right] J_1(\lambda), I_1(\lambda) \\ &= \left[T_{in} \sum a_i t_i(\lambda) \right] I_0(\lambda) + \left[r_i \sum a_i t_i^2(\lambda) \right] J_1(\lambda), J_1(\lambda) \\ &= r_g(\lambda) I_1(\lambda), \end{aligned}$$

from which is deduced the spectral reflectance $R(\lambda)$ of the halftone print:

$$R(\lambda) = \frac{J_0(\lambda)}{I_0(\lambda)} = r_s + \frac{T_{in} T_{ex} r_g(\lambda) [\sum a_i t_i(\lambda)]^2}{1 - r_i r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

This equation is the general Clapper-Yule equation. In their original paper, Clapper and Yule considered halftones of one ink, with surface coverage a and transmittance $t(\lambda)$:

Clapper-Yule Model,
Fig. 1 Transfers of light flux between the print-air interface and the diffusing substrate



$$R(\lambda) = r_s + \frac{T_{in}T_{ex}r_g(\lambda)[1 - a + at(\lambda)]^2}{1 - r_i r_g(\lambda)[1 - a + at^2(\lambda)]}.$$

If one Neugebauer primary covers the whole surface (“solid primary layer”), the equation becomes:

$$R(\lambda) = r_s + \frac{T_{in}T_{ex}r_g(\lambda)t^2(\lambda)}{1 - r_i r_g t^2(\lambda)},$$

and in the special case where no ink is printed (white colorant), it expresses the reflectance of the printing support:

$$R(\lambda) = r_s + \frac{T_{in}T_{ex}r_g(\lambda)}{1 - r_i r_g(\lambda)}.$$

Fresnel Terms and Measuring Geometries

The terms T_{in} , T_{ex} , r_s , and r_i are derived from the Fresnel formulae for unpolarized light. The Fresnel reflectivity of an interface between media 1 and 2, with respective indices n_1 and n_2 , is denoted as $R_{12}(\theta)$ when the light comes from medium 1 at the angle θ . The term T_{in} depends on the illumination geometry. It is typically:

$$T_{in} = 1 - R_{12}(\theta)$$

when directional light incomes at angle θ , or

$$T_{in} = 1 - \int_{\theta=0}^{\pi/2} R_{12}(\theta) \sin 2\theta \, d\theta$$

when the incident light is perfectly diffuse. T_{ex} depends on the measuring geometry. When the print is observed in one direction θ' , it is:

$$T_{ex} = (n_1/n_2)^2 T_{12}(\theta')$$

where the factor comes from the change of solid angle containing the exiting radiance due to the refraction. When an integrating sphere captures all light reflected by the print, it is:

$$T_{ex} = 1 - \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta \, d\theta.$$

The term r_s corresponds to the surface reflection, i.e., the gloss. It depends on the illumination and observation geometries, by it is typically 0.05 when it is captured, or 0 when it is discarded from measurement.

The reflectance r_i is the fraction of diffuse light emerging from the substrate which is reflected by the surface. It corresponds to the Lambertian reflectance of flat interfaces:

$$r_i = \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta \, d\theta.$$

For example, for a print made of material of refractive index 1.5 illuminated by directional light at 45° and observed at 0° (so-called $45^\circ/0^\circ$ measuring geometry), one has [2]:

$$\begin{aligned} T_{in} &= 1 - R_{12}(45^\circ) = 0.95, \\ T_{ex} &= T_{01}(0^\circ)/n^2 = \frac{1}{n^2} \left[1 - \frac{(1-n)^2}{(1+n)^2} \right] = 0.42, \\ r_s &= 0, \\ r_i &= 0.6. \end{aligned}$$

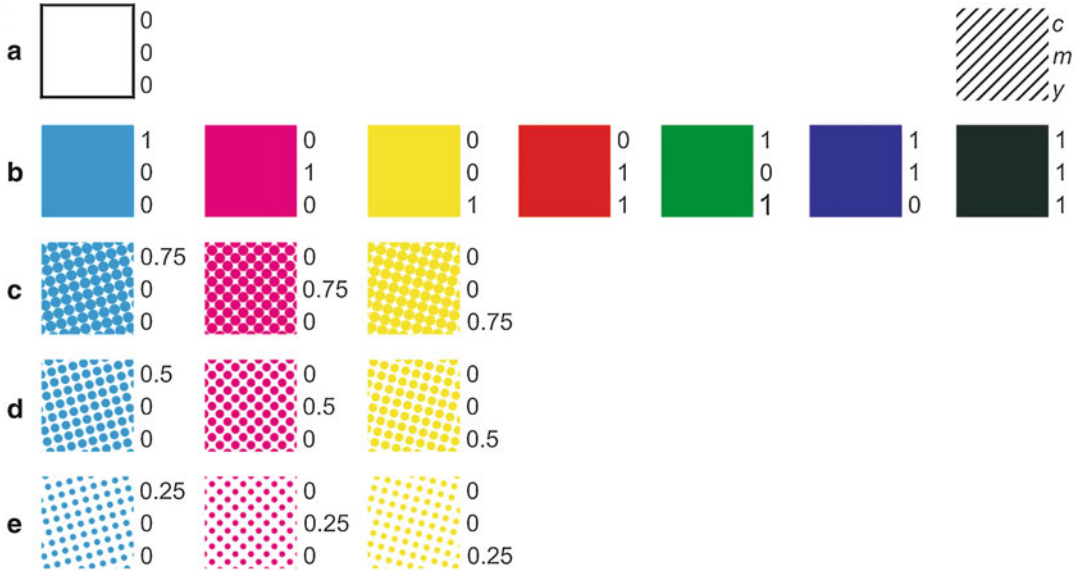
The Clapper-Yule equation becomes:

$$R(\lambda) = \frac{0.4r_g(\lambda)[\sum a_i t_i(\lambda)]^2}{1 - 0.6r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

Calibration of the Model and Prediction

The calibration of the model requires measuring the spectral reflectance of a few halftones, as the one represented in Fig. 2 in the case of CMY halftones.

By measuring the spectral reflectance $R_{w\text{left}}(\lambda)$ of the unprinted support (patch in row A of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:



Clapper-Yule Model, Fig. 2 Colors to be printed and measured using a spectrophotometer to calibrate the Clapper-Yule model in the case of CMY halftones

$$r_g(\lambda) = \frac{R_w(\lambda) - r_s}{T_{in}T_{ex} + r_i(R_w(\lambda) - r_s)}.$$

Then, by measuring the spectral reflectances $R_i(\lambda)$ of the solid primary patches (patches in row B of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:

$$t_i(\lambda) = \sqrt{\frac{R_i(\lambda) - r_s}{r_g(\lambda)[T_{in}T_{ex} + r_i(R_i(\lambda) - r_s)]}}.$$

All spectral parameters are now known. The spectral reflectance of a given halftone can be predicted by the Clapper-Yule general equation as soon as the surface coverages of different primaries are known. In classical clustered-dot or error diffusion prints, the primary surface coverages can be deduced from the surface coverages of the inks according to the Demichel equations. In the case of CMY halftones, the Demichel equations relating the surface coverages of the eight primaries to the surface coverages c , m , and y of the cyan, magenta, and yellow inks are:

$$\begin{aligned} a_w &= (1 - c) (1 - m) (1 - y), \\ a_c &= c(1 - m) (1 - y), \\ a_m &= (1 - c)m(1 - y), \\ a_y &= (1 - c) (1 - m)y, \\ a_{m+y} &= (1 - c)my, \\ a_{c+y} &= c(1 - m)y, \\ a_{c+m} &= cm(1 - y), \\ a_{c+m+y} &= cmy. \end{aligned}$$

Note that the prediction accuracy of the model sensibly depends on the exactitude of the primary surface coverage values. When the inks spread on the surface, which is almost always the case, the values for c , m , and y are larger than the ones defined in prepress. This phenomenon is generally called *dot gain*. The amount of ink spreading cannot be estimated by advance as it depends on several parameters, such as the chemical and mechanical properties of the inks and of printing support as well as the halftone pattern. They therefore need to be estimated from measurement.

Each ink i is printed alone on paper at the nominal surface coverages $q_i = 0.25, 0.5$, and 0.75 , which correspond to the nine color patches represented in rows C, D, and E of Fig. 2. Their

respective spectral reflectances are denoted as $R_{q_i}^{(m)}(\lambda)$. These halftones contain two primaries: the ink which should occupy a fractional area q_i and the paper white which should occupy the fractional area $1 - q_i$. Applying the Clapper-Yule equation with these two primaries and these surface coverages should predict a spectral reflectance $R_{q_i}^{(p)}(\lambda)$ equal to the measured one, but due to the fact that the effective ink surface coverage is different from the nominal one, these two reflectances are not the same. The effective surface coverage q_i' as the q_i value minimizing the deviation between predicted and measured spectra, by quantifying the deviation either by the sum of square differences of the components of the two spectra, i.e.,

$$q_i' = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left[R_{q_i}^{(p)}(\lambda) - R_{q_i}^{(m)}(\lambda) \right]^2$$

or by the sum of square difference of the components of their logarithm, i.e.,

$$q_i' = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left[\log R_{q_i}^{(p)}(\lambda) - \log R_{q_i}^{(m)}(\lambda) \right]^2$$

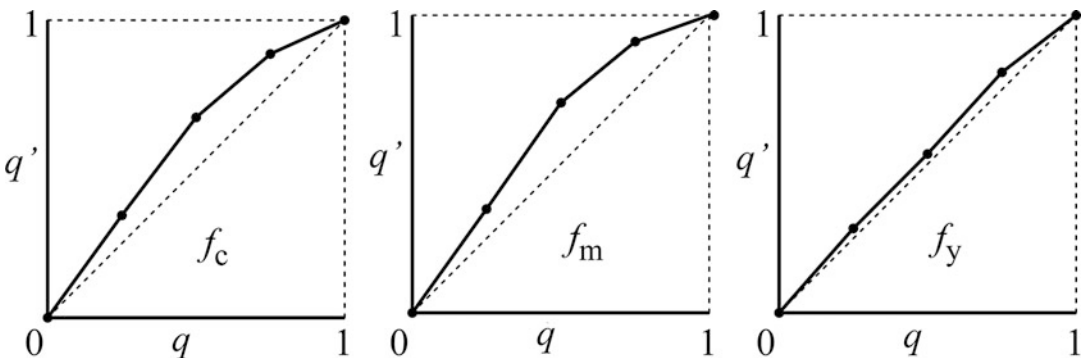
or by the corresponding color difference given, e.g., by the CIELAB metric

$$q_i' = \arg \min_{0 \leq q_i \leq 1} \Delta E_{94} \left(R_{q_i}^{(p)}(\lambda), R_{q_i}^{(m)}(\lambda) \right).$$

The first method is the most classical way of determining the effective surface coverage. Taking the log of the spectra as in the second method has the advantage of providing a higher weight to lower reflectance values where the visual system is more sensitive to small spectral differences. Fitting q_i' from the color difference metric sometimes improves the prediction accuracy of the model but complicates the optimization. Even at the optimal surface coverage q_i' , the difference between the two spectra is rarely zero and provides a first indication of the prediction accuracy achievable by the model for the corresponding print setup.

Once the nine effective surface coverages are computed, assuming that the effective surface coverage is 0, respectively 1, when the nominal surface coverage is 0 (no ink), respectively 1 (full coverage), three sets of q_i' values are obtained which, by linear interpolation, yield the continuous ink spreading functions f_i (Fig. 3). As an alternative, one can print halftones at nominal surface coverage 0.5 only and perform parabolic interpolation [3]. The number of patches needed for establishing the ink spreading curves is thus reduced to three (row C in Fig. 2).

Once the spectral parameters and the ink spreading functions are computed, the model is calibrated. The spectral reflectance of prints can be predicted for any nominal ink surface coverages c , m , and y . The ink spreading functions f_i



Clapper-Yule Model, Fig. 3 Example of ink spreading curves f obtained by linear interpolation of the effective surface coverages q_{ij}' deduced from the measured spectral

reflectances of patches with single-ink halftones (ink i) printed at nominal surface coverages 0.25, 0.5, and 0.75

directly provide the effective surface coverages c' , m' , and y' of the three inks:

$$\begin{aligned} c' &= f_c(c), \\ m' &= f_m(m), \\ y' &= f_y(y). \end{aligned}$$

Plugging these effective ink surface coverages into the Demichel equations, one obtains the eight effective primary surface coverages, and the general equation of the model finally predicts of the reflectance spectrum of the considered halftone.

Improved Ink Spreading Assessment Method

Hersch et al. observed that a given ink spreads differently according to whether it is printed alone on paper or superposed with another ink. They proposed an ink spreading assessment method taking into account the superposition conditions of the inks in the halftone [4]. This method increases noticeably the model's prediction accuracy. It relies on the halftones represented in Fig. 2 as well as the ones represented in Fig. 4.

The nominal ink surface coverages c , m , and y are converted into effective ink surface coverages c' , m' , and y' by accounting for the superposition-dependent ink spreading. In the case of CMY halftones, 12 in. spreading curves are obtained, similar to the ones represented in Fig. 5.

The effective surface coverage of each ink is obtained by a weighted average of the ink spreading curves, the weights being the surface coverages of the respective primaries on which the ink

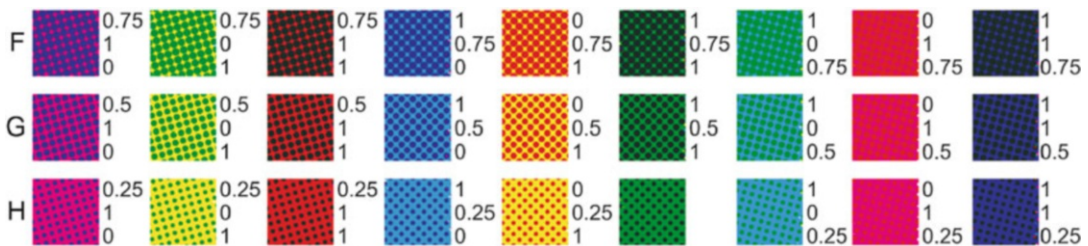
halftone is superposed. For example, the weight of the ink spreading curve f_c (cyan halftone over the white primary) is proportional to the surface of the underlying white primary, i.e., $(1 - m)(1 - y)$. First, $c' = c$, $m' = m$, and $y' = y$ are taken as initial values on the right side of the following equations:

$$\begin{aligned} c &= (1 - m)(1 - y)f_c(c_0) + m(1 - y)f_{c/m}(c_0) \\ &\quad + (1 - m)yf_{c/y}(c_0) + myf_{c/m+y}(c_0), \\ m &= (1 - c)(1 - y)f_m(m_0) + c(1 - y)f_{c/m}(m_0) \\ &\quad + (1 - c)yf_{m/y}(m_0) + cyf_{m/c+y}(m_0), \\ y &= (1 - c)(1 - m)f_y(y_0) + c(1 - m)f_{y/c}(y_0) \\ &\quad + (1 - c)m f_{y/m}(y_0) + cmf_{y/c+m}(y_0). \end{aligned}$$

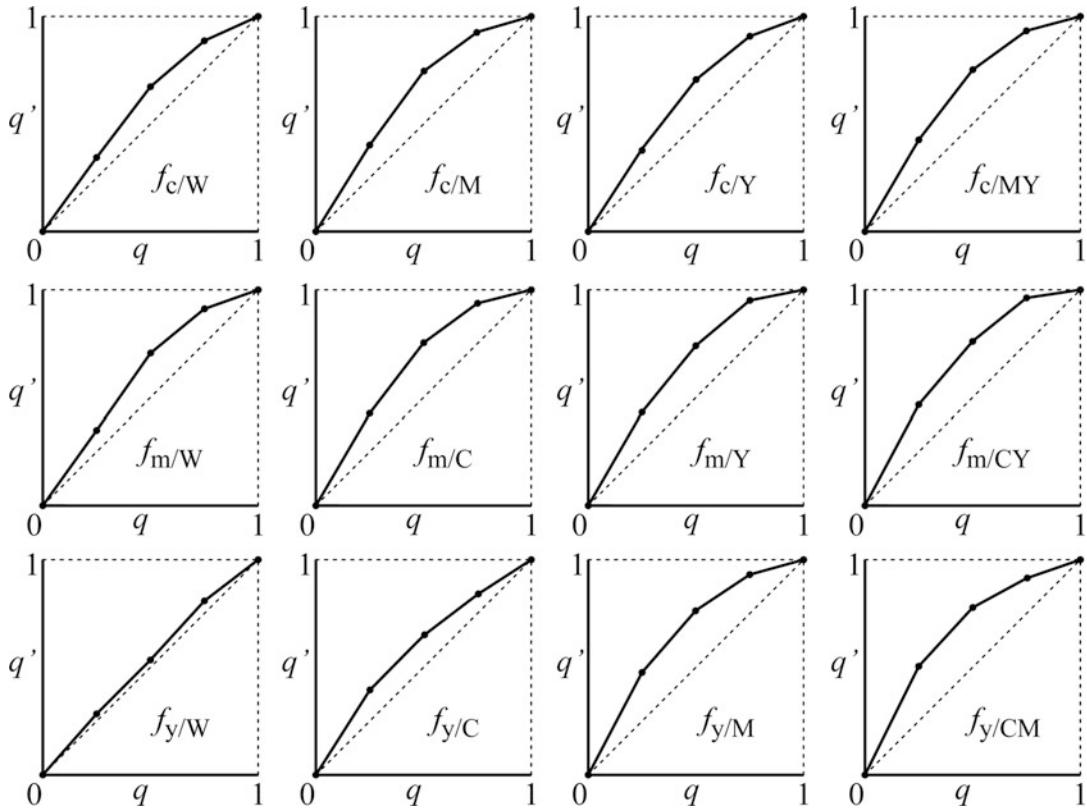
The obtained values of c' , m' , and y' are then inserted again into the right side of the equations, which yield new values of c' , m' , y' , and so on, until they stabilize. The final values of c' , m' and y' are plugged into the Demichel equations in order to obtain the effective surface coverages of the eight primaries. The spectral reflectance of the considered halftone is finally provided by the Clapper-Yule equation.

Experimental Testing

In order to assess the prediction accuracy of the model, predicted and measured spectra may be compared on sets of printed colors. As comparison metric, one generally uses the CIELAB ΔE_{94} , obtained by converting the predicted and measured spectra first into CIE-XYZ tristimulus values, calculated with a D65 illuminant and in respect to a 2° standard observer, and then into



Clapper-Yule Model, Fig. 4 Additional colors to be printed and measured using a spectrophotometer to assess the superposition-dependent ink spreading in the case of CMY halftones



Clapper-Yule Model, Fig. 5 Example of ink spreading curves f_{ij} for each ink i superposed with solid layers of the other inks in the case of CMY halftones

CIELAB color coordinates using as white reference the spectral reflectance of the unprinted paper illuminated with the D65 illuminant.

Because it assumes that the lateral propagation of light is large compared to the halftone screen period, the Clapper-Yule model is theoretically restricted to halftones with high screen frequency. For example, the model tested on two sets of 729 CMY colors printed with the same offset press on the same paper but at different frequencies, respectively, 76 and 152 lines per inch (lpi), provides better predictions for the highest frequency (average ΔE_{94} of 0.98 unit) than for the lowest one (average ΔE_{94} of 1.26 unit). Nevertheless, the experience shows that the model may also perform well for middle and low frequencies: For a set of 40 CMY colors printed in ink-jet at 90 lpi on supercalendered paper, the model

achieves a fairly good prediction accuracy, denoted by the average ΔE_{94} of 0.47 unit. Note that the average ΔE_{94} is 0.70 unit when the ink superposition conditions are not taken into account in the ink spreading assessment.

Conclusion

Despite the simplicity of its base equation, the Clapper-Yule model is one of the most accurate prediction models for halftone prints. Its main advantage compared to other models such as the Neugebauer model or the Yule-Nielsen-corrected Neugebauer model is the fact that physical parameters are attached to the different elements composing the print (inks, paper, and surface). The Fresnel terms can be adapted to the considered

measuring geometry, which is particularly interesting when predictions are made for a geometry different from the one used for calibration. The model also enables controlling ink thickness at printing time by comparing the colorant transmittances in various halftones, whose log is proportional to the ink thickness [5]. Recent improvements and extensions have been proposed which enable predicting both reflectance and transmittance of halftone prints thanks to extended flux transfer model relying on similar physical concepts as the Clapper-Yule model [6, 7].

Cross-References

- [Printing](#)
- [Reflectance Standards](#)

References

1. Clapper, F.R., Yule, J.A.C.: The effect of multiple internal reflections on the densities of halftone prints on paper. *J. Opt. Soc. Am.* **43**, 600–603 (1953)
2. Judd, D.B.: Fresnel reflection of diffusely incident light. *J. Res. Natl. Bur. Stand.* **29**, 329–332 (1942)
3. Rossier, R., Bugnon, T., Hersch, R.D.: Introducing ink spreading within the cellular Yule-Nielsen modified Neugebauer model. In: *IS&T 18th Color Imaging Conference*, San Antonio, TX, USA, pp. 295–300 (2010)
4. Hersch, R.D., Emmel, P., Crété, F., Collaud, F.: Spectral reflection and dot surface prediction models for color halftone prints. *J. Electron. Imaging.* **14**, 33001–33012 (2005)
5. Hersch, R.D., Brichon, M., Bugnon, T., Amrhyn, P., Crété, F., Mourad, S., Janser, H., Jiang, Y., Riepenhoff, M.: Deducing ink thickness variations by a spectral prediction model. *Color. Res. Appl.* **34**, 432–442 (2009)
6. Hébert, M., Hersch, R.D.: Reflectance and transmittance model for recto-verso halftone prints: spectral predictions with multi-ink halftones. *J. Opt. Soc. Am. A.* **26**, 356–364 (2009)
7. Mazaauric, S., Hébert, M., Simonot, L., Fournel, T.: Two-flux transfer matrix model for predicting the reflectance and transmittance of duplex halftone prints. *J. Opt. Soc. Am. A.* **31**, 2775–2788 (2014)

Clinical Color Vision Tests

- [Color Vision Testing](#)

Coarseness

- [Texture Measurement, Modeling, and Computer Graphics](#)

Coatings

- [Inorganic, Hybrid, and Functional Pigments](#)

Coefficient of Utilization, Lumen Method

Peter Thorns

Strategic Lighting Applications, Thorn Lighting Ltd, Durham, UK

Definition

The lumen method is an indoor lighting calculation methodology that allows a quick assessment of the number of luminaires necessary to achieve a given average illuminance level or alternatively the average illuminance level that will be achieved for a given number of luminaires. It is valid for empty rectangular rooms with simple three-surface diffuse reflectances for ceiling, wall, and floor.

Introduction

When an electric lamp is turned on, it emits light and it is possible to quantify the amount of light by measuring it, the result being given in the units of lumens. In itself this is a useful piece of information, but what would be more useful would be to have a method of converting this into a measure of the amount of light that would be received onto a desk from one or more luminaires or alternatively the number of luminaires necessary to achieve a given quantity of light on the desk. This calculation is known as the lumen method.

Light falling onto a surface is called illuminance and has units of lux (lx) or lumens per square meter (lm/m^2). It is not possible to see illuminance as light is actually invisible, which is fortunate, else any view would be degraded by looking through a fog of light. What is actually seen is the effect light has on surfaces, the reflected light, and this is called luminance with units of candelas per square meter (cd/m^2). To test this, place a sheet of white paper onto a dark surface (see Fig. 1). At the point where the edge of the paper meets the dark surface, the amount of light falling onto the two materials will be approximately the same. However, they appear completely different as the eye detects the light reflected back towards it, not the light falling onto the surfaces, and the white paper reflects more light than the dark desktop.

However, luminance is a difficult quantity to measure and changes with viewing position. Imagine a day-lit room with resultant shadows and patches of light reflected from polished surfaces. As the observer position moves within the space, the shadows and patches of light change with viewing position. As vision is essentially viewing luminance, this means that the luminance is view dependent. This makes it a difficult design quantity, so for ease, common practice is to design using illuminance which is view independent as

the amount of light falling onto a surface does not change with viewing position. (There are a few exceptions, an example being traffic routes, where design may use luminance, in this case the light that is reflected from the road surface).

Calculating the Maximum Achievable Illuminance

To be able to calculate the amount of light falling onto a surface from a given number of luminaires, it is necessary to be able to convert between the measures of lamp lumens (lm) and illuminance (lx). Remember lx is also lm/m^2 , so a given total quantity of lumens can be converted into illuminance by dividing it by the size of the area to be lit in m^2 . But it is important to consider where the area to be lit really is. Ideally luminaires should provide task lighting, that is, light the task being undertaken at any given point to the correct level of illuminance. This means that if there is a desk in a room, it would generally be lit to a level of 500 lx, the recommended level of illuminance for reading and writing tasks [4]. However, the circulation space within the room does not need this quantity of light, 200 lx being a perfectly adequate illuminance level to safely move around the space. This has two problems:

- Frequently in large spaces, it is not known where tasks such as desks will be positioned and may not be even clearly known what tasks will be performed within the space. Even if the task types and locations are known, most spaces and how they are used change through time. This means that, within reason, lighting has to be flexible enough to preserve the correct lighting conditions for changing requirements.
- Calculating the illuminance level for a particular area within a larger space is generally beyond the ability of a quick and easy calculation method. For lighting the calculation of an average illuminance across a space is relatively easy, while the calculation of the illuminance across a desk located in a room would normally



Coefficient of Utilization, Lumen Method,
Fig. 1 White paper on a black desk

require the use of a computer calculation program.

Therefore, to keep the calculation simple, it will calculate the average illuminance across a horizontal plane, called the task plane. This could either be the floor or a virtual plane at desk height depending upon the expected height of the task. So remembering the units of lm/m^2 , it is necessary to know the total quantity of lumens within the space. So

- For a given number of luminaires, N_{lum}
- Where each luminaire contains a given number of lamps, N_{lamp}
- And each lamp produces a given quantity of lumens, lm_{lamp}

the total amount of lumens available, lm_{total} , is equal to

$$\text{lm}_{\text{total}} = N_{\text{lum}} * N_{\text{lamp}} * \text{lm}_{\text{lamp}} \quad (1)$$

So for a given area of task plane, A_{tp} , and from Eq. 1, the known total amount of lumens available, lm_{total} , the absolute maximum illuminance possible, lx_{max} , will be

$$\text{lx}_{\text{max}} = \text{lm}_{\text{total}} / A_{\text{tp}} \quad (2)$$

There are two points to consider regarding the value produced by Eq. 2.

- This value assumes *all* the light produced by the lamps will be received onto the task, with no losses. This is generally not possible or even desirable.
- As this is the maximum illuminance theoretically (but not practically) achievable, if the illuminance is too low, at this point more luminaires and, hence more lumens, will be required.

So how is this result converted into a more practical value? The losses in light come from three main effects:

- Losses within the light fitting
- Losses through aging and dirt

- Losses through light not going directly to the task plane but via reflection from room surfaces

Accounting for Losses Within the Light Fitting

A luminaire is a housing containing a light source with associated control gear, optics, wiring, and electrical connections, both to the light source and to the external power supply. This creates a number of problems:

- The luminaire has its own microclimate, the temperature inside the fitting generally being different to that of the air outside. Designing a fitting so that the internal temperature is the optimum for the light source it contains requires skilled design, and the larger the difference between the optimum light source conditions and the actual luminaire conditions results in an increasing reduction in lumen output from the light source and/or reduced operating life expectancy for the source.
- An optic, no matter how well designed, has an element of inefficiency. No surface will reflect 100% of any light incident upon it and losses are cumulative. So, for example, if a metal reflector has a reflectance of 92%, then light that bounces once off the reflector before exiting the luminaire loses 8% of its initial lumens. If it requires two bounces to exit the fitting, only $0.92 * 0.92 = 0.85$ or 85% of the initial lumens exit the fitting, 15% being lost. Similarly optics that rely on transmission of light, such as diffusers or prismatic controllers, absorb some of the light, no material being 100% transmissive.
- Components within the luminaire will obstruct light, and this light may become trapped and never exit the fixture. For example, light emitted from the light source upwards into a ceiling recessed luminaire will need to be reflected around the light source, and some light will be lost in this process, trapped behind the source. (This can cause extra problems if the light is absorbed by the source, causing it to heat up further). Many reflectors trap light in

their back which may be open and shaped as the inside of a V with low reflectance.

Therefore, a measure is needed to quantify how all of these situations affect the lumen output of the luminaire. This is the light output ratio (LOR). Essentially the LOR is the ratio of light emitted from the luminaire to the light emitted by the light source, and the value of LOR is luminaire specific.

So to account for these effects, Eq. 2 is modified as shown in Eq. 3 below:

$$I_{x_{\text{luminaires}}} = I_{m_{\text{total}}} * \text{LOR} / A_{tp} \quad (3)$$

Accounting for Losses Through Aging and Dirt

When a lighting system is first installed, the lamps are new and all functioning, the luminaires are clean, and generally the room surfaces (floor, walls, and ceiling) are clean. However, through time the condition of the installation will deteriorate. As light sources age, their lumen output reduces (lumen depreciation) and some lamps will fail completely. Dust and dirt will gather on the reflecting surfaces of the luminaires, reducing how efficiently light is directed from the light source out of the fitting, and room surfaces will

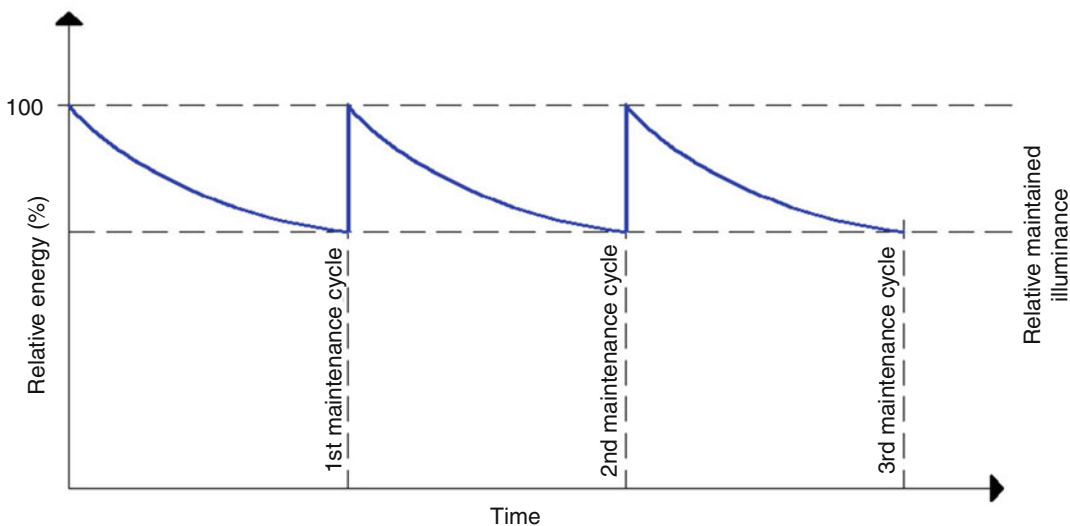
become dirty and marked through everyday wear and tear. All of these will reduce the amount of light reaching the task plane.

A suitable maintenance routine, such as renewing aging lamps and cleaning the luminaires, will help minimize these impacts, but the amount of light received onto the task plane will still vary through life, reducing through time between maintenance cycles, increasing back to close to the original lighting levels immediately after a maintenance cycle (see Fig. 2).

However, a lighting design should be designed to produce a level of maintained illuminance, so that even at the point of maximum reduction in light just before maintenance is performed, the required light level is still achieved. To account for this, a maintenance factor (mf) is used, which is the amount of light lost when the light sources are at their oldest and the luminaires and room surfaces are at their dirtiest. So modifying Eq. 3 produces

$$I_{x_{mf}} = I_{m_{\text{total}}} * \text{LOR} * mf / A_{tp} \quad (4)$$

Further advice on the determination of maintenance factors is available from the Commission Internationale de L'Eclairage [3].



Coefficient of Utilization, Lumen Method, Fig. 2 A typical scheme maintenance cycle

Accounting for Reflection Losses from Room Surfaces

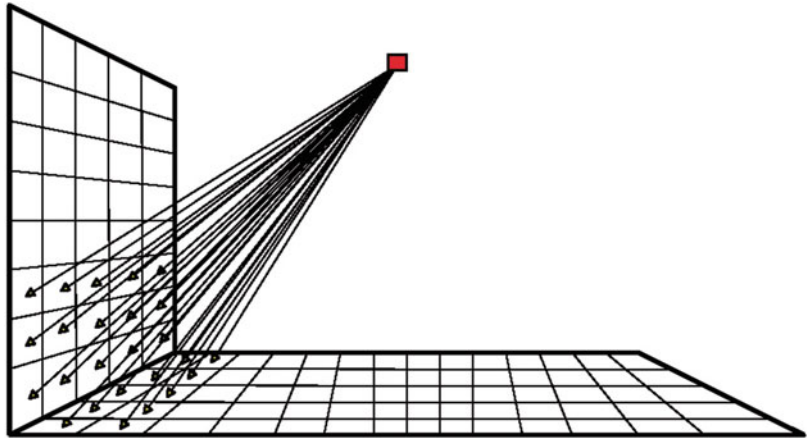
Equation 4 still makes one major assumption that all of the light from the luminaires goes directly onto the working plane.

However, this is rarely the case, light being directed onto the room surfaces (and it is rarely desirable for all light to go directly to the task as this would result in a pool of light within a dark room which would be an uncomfortable work environment). So some of the light hits the room surfaces (see Fig. 3), and some of this light will be reflected back onto the task (see Fig. 4). However, it should be remembered that no surface will have 100% reflectance. The quantity of light reflected will depend upon the material properties of a

surface, a light-colored wall typically having a reflectance in the region of 60%, so a quantity of light will be lost with each reflection from a surface. To account for this, a measure called the *utilization factor* (UF) is used. This is a measure of the total amount of light from a luminaire that reaches the task, both directly from the luminaire and indirectly through reflection from room surfaces. Manufacturers publish tables of utilization factors, and these vary as values are dependent upon the luminaire distribution, the room reflectance, and also upon the ratio of wall surface area to ceiling/floor surface area. (Given a ceiling with a high reflectance within a room that is tall and thin proportionally, there is a large amount of wall surface area to ceiling surface area, so the high reflectance ceiling will have less effect than for a

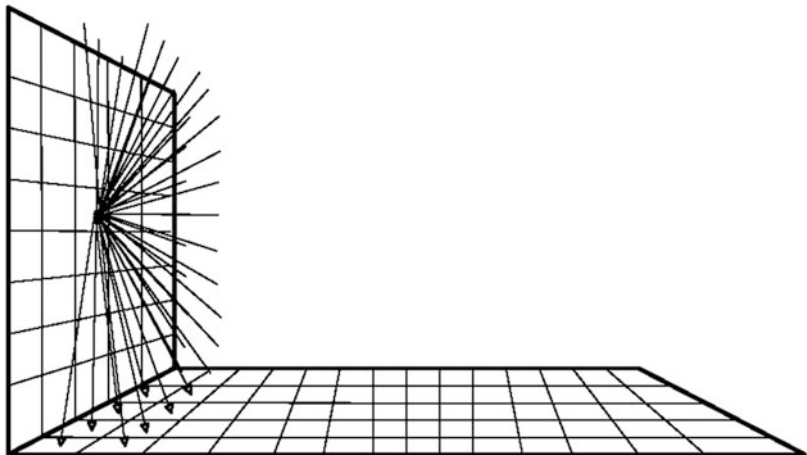
Coefficient of Utilization, Lumen Method,

Fig. 3 Light distribution onto the wall and floor



Coefficient of Utilization, Lumen Method,

Fig. 4 Light distribution from the wall to other room surfaces



large open plan office with a large ceiling surface area compared with the wall surface area).

To calculate the ratio of wall to ceiling/floor surface area, the room index (K) is used. This is defined as

$$K = (\text{area of ceiling} + \text{area of floor}) / \text{total area of wall}$$
(5)

For a room of length L, width W, and height H (where the height is the distance between the task plane and the luminaire plane, see Fig. 5), Eq. 5 becomes

$$K = ((L * W) + (L * W)) / (2(L * H) + 2(W * H))$$

Therefore,

$$K = 2(L * W) / 2H(L + W)$$

Therefore,

$$K = L * W / (L + W)H$$
(6)

So given the example table of utilization factors shown in Table 1, it can be seen that values of

utilization factor are supplied for a variety of room indices and surface reflectance's. From the table, it can be seen that for a room with reflectance's of

- Ceiling 70%
- Walls 50%
- Floor 20%

and with dimensions

- Length 12 m
- Width 12 m
- Height 3 m

which give a room index of

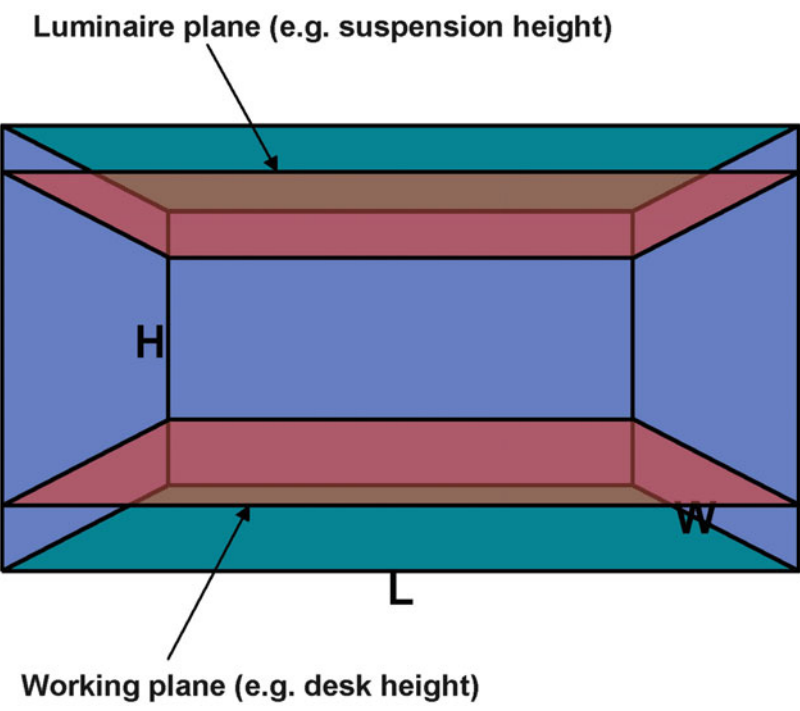
$$K = (12 * 12) / ((12 + 12) * 3) = 2$$

the utilization factor would be 0.69.

(Values used in the calculation are always percentages, so in this case, 69% = 0.69. Some tables may already show the values as fractions, in this case 0.69).

Therefore, using Eq. 4 adjusted by the utilization factor gives

Coefficient of Utilization, Lumen Method, Fig. 5 Room dimensions in terms of L, W, and H



Coefficient of Utilization, Lumen Method, Table 1 An example utilization factor table

Utilization Factors								
Room Reflectance	Room Index							
Ceiling/Walls/Floor	0.75	1.00	1.25	1.50	2.00	2.50	3.00	4.00
70/50/20	50	56	61	64	69	72	74	77
70/30/20	44	51	56	60	65	68	71	74
70/10/20	41	47	52	56	62	65	68	72
50/50/20	48	54	59	62	66	69	71	74
50/30/20	44	50	55	58	63	66	69	72
50/10/20	41	47	52	55	61	64	66	70
30/50/20	47	53	57	60	64	67	69	71
30/30/20	43	49	54	57	62	65	67	69
30/10/20	40	46	51	55	59	63	65	68
0/0/0	39	44	49	53	57	60	62	65

$$lx_{final} = lm_{total} * LOR * mf * UF / A_{tp} \quad (7)$$

So with knowledge of the type and number of luminaires in a space, it is possible to calculate approximately how much illumination the space will have.

Information on the calculation of utilization factor tables is available from the Commission Internationale de L'Eclairage [1, 2].

Calculating the Number of Luminaires Required in a Room

Equation 7 may also be rearranged to allow the calculation of the required number of luminaires to achieve a given illumination level.

$$lm_{total} = lx_{final} * A_{tp} / LOR * mf * UF \quad (8)$$

And using Eq. 1 gives

$$N_{lum} * N_{lamp} * lm_{lamp} = lx_{final} * A_{tp} / LOR * mf * UF$$

So the number of luminaires required to achieve a given level of illumination is

$$N_{lum} = lx_{final} * A_{tp} / LOR * mf * UF * N_{lamp} * lm_{lamp} \quad (9)$$

Cross-References

- [Interior Lighting](#)
- [Light Distribution](#)
- [Lumen Depreciation](#)
- [Luminaires](#)

References

1. CIE Publication 40: Calculations for interior lighting – basic method (1978)
2. CIE Publication 52: Calculations for interior lighting – applied method (1982)
3. CIE Publication 97: Guide on the maintenance of indoor electric lighting systems (2005)
4. ISO 8995-1:2002(E)/CIE S 008/E: 2001: lighting of work places part 1: indoor

Coherence

- [Color Harmony](#)

Cold Body Radiation

- [Luminescence](#)

Cold Light

- [Luminescence](#)

Collective Intelligence

- [Crowdsourcing of Multilingual Color Names](#)

Color Accord

- [Color Harmony](#)

Color Adaptation

- [Environmental Influences on Color Vision](#)

Color Aesthetics

- [Color Preference](#)

Color and Culture

- [Environmental Influences on Color Vision](#)

Color and Light Effects

- [Interactions of Color and Light](#)

Color and Light Interdependence

- [Interactions of Color and Light](#)

Color and Lightness Constancy

- [Retinex Theory](#)

Color and Visual Search, Color Singletons

Jasna Martinovic¹ and Amanda Hardman²

¹Department of Psychology, University of Edinburgh, Edinburgh, UK

²School of Psychology, University of Aberdeen, Aberdeen, UK

Synonyms

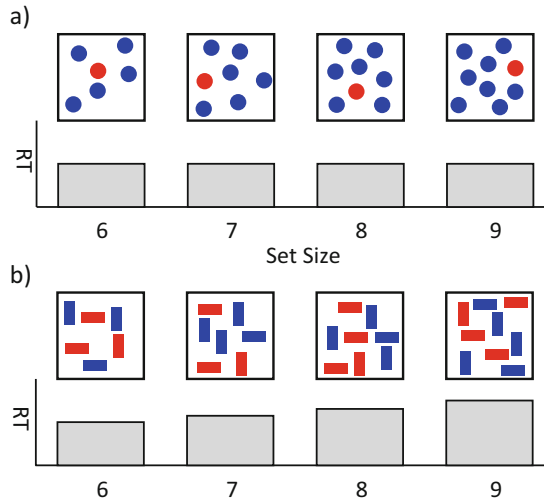
[Color pop-out](#)

Definition

Visual search is a task requiring detection of one or more target items within a multi-item display populated by both target(s) and distractors.

Characteristics of the Visual Search Task

In a visual search task, the item that is being searched for is known as the **target**; other items are known as **distractors**. Figure 1 presents an example of visual search displays with one target item and a varying number of distractor items. Participant's task is to indicate whether a target is present or absent in the display, or to report an attribute of the target (e.g., whether it is on the left or right side). The total number of items in the display is known as the **set size**. Items in the search display can differ along various **feature dimensions**, for example, color or orientation. Individual items can take on different **feature values** from that dimension – for example, color (red or blue), or orientation (vertical or horizontal). If items in the visual search task vary along a



Color and Visual Search, Color Singletons, Fig. 1 Examples of visual search displays. A range of different set sizes is shown, going from 6 to 9 items. **(a)** Feature search: Target is a red circle. **(b)** Conjunction search: Target is a red vertical bar. Reaction time for each

set size is shown underneath each search set. Reaction times for feature search are typically independent of the set size, while reaction times for conjunction search most often increase linearly with the addition of extra items

single dimension, such as color, the observer is commonly asked to look for a target of a specified hue (e.g., red) in a set of distractors that differ in hue. This is known as **feature search**. In the case of feature search, distractors can be heterogeneous – all varying in hue – or homogeneous – all of them being of the same hue (e.g., blue, as in the example in Fig. 1a). If all distractors share the same color, the uniquely-colored target is defined as a **target singleton**. Note that **irrelevant singletons** are sometimes also used in visual search tasks, e.g., a single orange distractor can be introduced into a display with a red target and a number of otherwise blue distractors. Detection of color-singleton targets is very efficient, as long as there are no salient task irrelevant singletons. Such reaction times usually are independent of set size (see Fig. 1a). This is known as the **pop-out** effect characteristic of **parallel visual search**. If, however, items vary along multiple dimensions, the participant is commonly asked to look for, e.g., a red bar oriented upright among a set of distractors that differ in both hue and orientation (see Fig. 1b). This is known as **conjunction search**. Conjunction search is typically less efficient, producing reaction time costs as set size

increases, which is a characteristic of **serial visual search**.

Visual search: A Historical Overview

Visual search is one of the most commonly used tasks in attention research. The popularity of the visual search task stems from the fact that it operationalizes a vital activity performed by both humans and non-human animals. Eckstein's review [4] summarizes many examples of everyday search situations. In natural environments, foraging for food involves searching for edible fruit, while in man-made environments, operators monitor complex images in order to detect security-relevant or medically-relevant information. In many real world search situations, color is an important determinant of performance due to its ability to make certain features of the scene more or less conspicuous. For example, in order to avoid detection by predators, prey often adopts coloration that acts as camouflage, ensuring that it does not “pop-out” when seen in its natural environment.

Visual search became a prominent laboratory task for studying attention in the 1980s providing the initial evidence base for Treisman's highly influential Feature Integration Theory (FIT). FIT posits that attentional deployment is guided by multiple, distinct feature maps that can be activated in parallel [13]. Visual search represents an ideal task to test this theory, with the potential to reveal underlying neural representations of feature maps by using relatively simple, behavioral methods (for a review, see [10]). The slope of the reaction time (RT) function (milliseconds per item) was considered to be a particularly informative variable that provides insight into the amount of time required for an observer to process one item before moving on to the next item. In parallel search, the RT function slope is constant across various set sizes, which, according to FIT, is due to the target's unique representation in a retinotopic feature map activated in parallel to other basic-level maps. Color was shown to be one of the basic, pre-attentive features for visual search implying that it can be identified in parallel (i.e., without the deployment of serial attention) across the entire visual field. One of the most notable models of visual search to follow FIT is Wolfe's guided search model (for details, see [14]).

Determinants of Visual Search for Color

Color is considered one of the basic visual features since it can support parallel visual search. In fact, color has repeatedly been shown to be one of the most potent feature dimensions for guiding visual search (e.g., [1]). As long as the difference in chromaticity between the target and the distractors is sufficiently large, search for color is highly efficient and response times remain constant regardless of display density [9].

Clearly, proximity of items' chromaticity in color space is the key determinant of color search, with increased efficiency for targets that are closer to each other and further away from the distractors ([11]; see also [8] for converging evidence using a different experimental task). But what else affects search for color? Perhaps some of the low-level,

physiological signals are more salient than others? Alternatively, search could be more efficient for targets that are close to category prototypes. To answer this question, Lindsey et al. [7] contrasted contributions from cone-opponent, color-opponent or categorical representations when searching for a desaturated color singleton. They found that cone-opponent chromatic representations determined the speed of visual search, with bluish colors being particularly disadvantaged. At first, it may be surprising that low-level, cone-opponent representations enacted a more powerful influence on color search than higher-level, color-opponent or categorical representations. The two cone-opponent mechanisms can be mapped onto the two physiological pathways that extract chromatic signals: one distinguishes between reddish and greenish hues through opposing signals from L and M cones (L-M mechanism), while the other distinguishes between bluish and yellowish hues through opposing the S-cone signal and a combination of L and M cone signals (S-(L + M) mechanism). Lindsey et al. found that search was particularly ineffective when searching for desaturated (i.e., lighter) bluish colors, which were aligned with the bluish pole of the S-(L + M) mechanism. Meanwhile, search was particularly effective for pinkish colors, which fell in proximity of the reddish pole of the L-M mechanism. However, while the study by Lindsey and colleagues strongly suggests that cone-opponent signals are critical in driving search reaction times, the relation of such cone-opponent effects to the level of luminance contrast in the stimulus remains somewhat unclear. Li et al. [6] demonstrated that while search reaction times for targets defined by L-M contrast decrease after luminance signals are added, this is not the case for targets defined by S-cone contrast. Asymmetrical interactions between luminance and chromatic signals in determining target salience, especially as reflected in response speed, could provide a mechanism through which cone-opponent signals are able to influence visual search performance when searching for a desaturated target. However, relying solely on the visual search task to study representations that underlie attention to color may

be problematic as the task itself combines bottom-up, i.e., salience-driven, and top-down, i.e., goal-driven, influences on attention. A recent study by Ansorge and Becker [2] used a spatial cueing paradigm instead of the classical visual search in order to circumvent that problem, but the results failed to support a single representational space being employed for color selection.

During visual search, target features are maintained in working memory. This implies that visual search for color necessarily shares some constraints with working memory for color. Representations in working memory are impacted by color categories: they are significantly biased away from category boundaries and toward category centers [3]. These higher-level, category-driven computations may enact an important influence on search templates. In a similar vein, if targets and distractors can be distinguished on the basis of a relational criterion, e.g., “redder than” or “yellowier than,” absolute featural tuning gets overruled by such relational tuning. For example, in a study by Harris et al. [5] observers searched for orange among yellow distractors by selecting items that were “more reddish” when trials were blocked together. They tuned into orange as the particular feature only when the search displays of orange singletons among red distractors were randomly interleaved with search displays of orange singletons among yellow distractors, rendering such category-related, relational search templates ineffective.

Concluding Comments

Visual search experiments led to important discoveries on attentional selection of color. While visual search remains a highly useful task, it may be advantageous to consider the knowledge gained from visual search in a broader context: in recent years, foraging tasks yielded information on how we search for multiple targets that belong to different categories [15], while statistical summary representation tasks have been successfully used to measure attentional tuning into color [12].

Cross-References

- [Categorical Perception](#)
- [Psychological Color Space and Color Terms](#)

References

1. Alexander, R.G., Nahvi, R.J., Zelinsky, G.J.: Specifying the precision of guiding features for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **45**(9), 1248–1264 (2019). <https://doi.org/10.1037/xhp0000668>
2. Ansorge, U., Becker, S.I.: Contingent capture in cueing: The role of color search templates and cue-target color relations. *Psychol. Res.-Psychol. Forsch.* **78**(2), 209–221 (2014). <https://doi.org/10.1007/s00426-013-0497-5>
3. Bae, G.Y., Olkkonen, M., Allred, S.R., Flombaum, J.I.: Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *J. Exp. Psychol.-Gen.* **144**(4), 744–763 (2015). <https://doi.org/10.1037/xge0000076>
4. Eckstein, M.P.: Visual search: A retrospective. *J. Vis.* **11**, 5 (2011)
5. Harris, A.M., Remington, R.W., Becker, S.I.: Feature specificity in attentional capture by size and color. *J. Vis.* **13**(3) (2013)
6. Li, J.C., Sampson, G.P., Vidyasagar, T.R.: Interactions between luminance and colour channels in visual search and their relationship to parallel neural channels in vision. *Exp. Brain Res.* **176**(3), 510–518 (2007). <https://doi.org/10.1007/s00221-006-0773-3>
7. Lindsey, D.T., Brown, A.M., Reijnen, E., Rich, A.N., Kuzmova, Y.I., Wolfe, J.M.: Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychol. Sci.* **21**(9), 1208–1214 (2010). <https://doi.org/10.1177/0956797610379861>
8. Martinovic, J., Wuerger, S.M., Hillyard, S.A., Müller, M.M., Andersen, S.K.: Neural mechanisms of divided feature-selective attention to colour. *NeuroImage.* **181**, 670–682 (2018). <https://doi.org/10.1016/j.neuroimage.2018.07.033>
9. Nagy, A.L., Sanchez, R.R.: Critical color differences determined with a visual-search task. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **7**(7), 1209–1217 (1990). <https://doi.org/10.1364/josaa.7.001209>
10. Nakayama, K., Martini, P.: Situating visual search. *Vis. Res.* **51**(13), 1526–1537 (2011). <https://doi.org/10.1016/j.visres.2010.09.003>
11. Stroud, M.J., Menneer, T., Cave, K.R., Donnelly, N.: Using the dual-target cost to explore the nature of

- search target representations. *J. Exp. Psychol.-Hum. Percept. Perform.* **38**(1), 113–122 (2012). <https://doi.org/10.1037/a0025887>
12. Sun, P., Chubb, C., Wright, C.E., Sperling, G.: Human attention filters for single colors. *Proc. Natl. Acad. Sci. U. S. A.* **113**(43), E6712–E6720 (2016). <https://doi.org/10.1073/pnas.1614062113>
 13. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980)
 14. Wolfe, J.M.: Visual search: How do we find what we are looking for? *Annu. Rev. Vis. Sci.* **6**(1), 539–562 (2020). <https://doi.org/10.1146/annurev-vision-091718-015048>
 15. Wolfe, J.M., Cain, M.S., Aizenman, A.M.: Guidance and selection history in hybrid foraging visual search. *Atten. Percept. Psychophys.* **81**(3), 637–653 (2019). <https://doi.org/10.3758/s13414-018-01649-5>

Color Appearance

- [Appearance](#)

Color Appearance Solid

- [Psychological Color Space and Color Terms](#)

Color Atlases

- [Color Order Systems](#)

Color Bleeding

- [Global Illumination](#)

Color Blindness

- [Deuteranopia](#)
- [Protanopia](#)
- [Tritanopia](#)

Color Board

- [Palette](#)

Color Categorical Perception

J. Richard Hanley

Department of Psychology, University of Essex,
Colchester, UK

Definition

Categorical perception (CP) occurs when discrimination of items that cross category boundaries is faster or more accurate than discrimination of exemplars from the same category. Categorical perception of color is observed when, for example, a green stimulus and a blue stimulus are more easily distinguished than two stimuli from the same color category (e.g., two different shades of green). Color differences in terms of discriminability can be equated across between-category and within-category comparisons by using the *Commission Internationale de L'Eclairage* (CIE) values. It is therefore important to note that superior cross-category relative to within-category discrimination is observed when the physical distance between cross-category items and the physical distance between within-category items are equivalent.

Categorical Perception Using a Two Alternative Forced-Choice Procedure

Categorical perception of color has been demonstrated many times using a two alternative forced-choice (2-AFC) procedure. In an experiment of this kind, participants view a colored patch for a short duration. Shortly afterward, the original item is displayed next to a distractor, and the participant has to indicate which of the two colored patches was presented a few moments earlier. Discrimination between the target and distractor

is significantly more accurate when they belong to different color categories (e.g., green target and blue distractor) than when they come from within the same color category (e.g., target and distractor are different shades of green).

Categorical perception on the 2-AFC task has provided evidence for the existence of different color categories in different cultures [1]. For example, the Berinmo are a traditional hunter-gatherer culture in the upper Sepik region of Papua New Guinea who have a different set of basic color terms from speakers of English. Berinmo speakers showed significantly better discrimination of 32 cross-category items than 32 within-category items at the boundary between two Berinmo color categories (*nol* and *wor*) that do not exist for English speakers [1]. Conversely, English speakers did not show CP at this boundary. Most important, there was no evidence of CP at the boundary between green and blue for speakers of Berinmo whose language does not make this distinction. Such findings provide evidence against the idea that color categories are universal and exist irrespective of the color vocabulary that speakers have acquired during language development.

Hanley and Roberson [2] argued that CP in the 2-AFC task reflects the role of categorical codes in distinguishing targets from distractors. They proposed that, in principle, participants can perform the 2-AFC task using either categorical or perceptual information about the target item. The categorical code contains less information but may generally be easier to retain than the perceptual code. Targets and distractors can be more accurately distinguished when they cross a category boundary because they differ at both the categorical and the perceptual level. The categorical code cannot be used to distinguish targets and distractors from the same color category, however, so participants must rely entirely on the perceptual code. This account can readily explain the finding that CP for color was abolished when a verbal interference task was interposed between presentation of the target and the test pair [3]. Presumably, verbal interference disrupts participants' ability to generate or retain the category code, and so performance for both within- and

cross-category targets relied on retention of a perceptual representation of the color of the target.

Hanley and Roberson [2] also reanalyzed the results of a series of studies that had demonstrated CP effects for color using the 2-AFC task. This reanalysis revealed that within-category performance was particularly weak when the target was a poor example of a color category and the distractor was a better exemplar. If, however, the target was a good exemplar and the distractor a poor exemplar, performance on within-category trials and cross-category trials was equivalent. Such an outcome is of considerable theoretical importance because it challenges the commonly held view that CP has a perceptual basis (e.g., [4]). According to this account, CP occurs because perceptual systems are fine-tuned by experience such that sensitivity to change is greater around category boundaries than it is to changes that occur within a category. Exactly the same perceptual discriminations are required for within-category trials in which performance is good (central target, peripheral distractor) as for those in which performance is poor (peripheral target, central distractor). Consequently, the advantage for central targets over peripheral targets, which forms the basis for CP in the 2-AFC task, cannot reflect greater perceptual sensitivity for central targets.

Why is performance so poor on the 2-AFC task when targets are peripheral examples of a color category? Hanley and Roberson [2] argued that the poor exemplars in these studies were less likely to be classified as category members when they appeared at test alongside a good exemplar than when they appeared by themselves at presentation. The likely outcome is a mismatch between the way in which a peripheral target is categorized when it is originally presented and the way it is classified at test. A mismatch of this kind will lead to poor performance on the 2-AFC task because the categorical code now provides misleading information about the identity of the target color. This explanation is consistent with the findings that classification of an exemplar at the center of a color category was unaffected by the presence of a peripheral category member [5]. But, crucially, when a poor exemplar appeared alongside a

central exemplar, the poor exemplar was frequently assigned a different category label from the one it was given when it appeared on its own. Similar effects have been observed in other domains and have been termed “category contrast effects” [6].

Consistency of categorization is crucial to the account of CP for color in the 2-AFC task suggested by Hanley and Roberson [2]. For within-category pairs, it might appear that a categorical code could not possibly distinguish target from distractor. However, it follows from the category contrast effect that when the target is a central exemplar and the distractor is a poor exemplar, the distractor is likely to be given a different categorical label from the target. Hence, Hanley and Roberson argued that good performance in the 2-AFC task when the target is a central exemplar occurs because target and distractor can often be accurately distinguished on the basis of the categorical code. It also follows from the category contrast effect that performance will be poor on within-category trials when the target is a peripheral exemplar. For example, a greenish-blue boundary target may be categorized as “blue” when it appears on its own. However, the presence at test of a distractor that is a good example of “blue” means that the peripheral “blue” target will sometimes elicit a different category label (“green”) at test. The outcome will be that the category label given to the target at encoding is elicited at test only by the distractor, which is likely to be selected in preference to the target as a consequence.

Categorical Perception in Visual Search Tasks

Even though test stimuli are presented relatively soon after the target has been removed, performance on the 2-AFC task necessarily depends on a memory code for the target color. Recent studies have also found CP for color in visual processing tasks where the memory component is minimized. For example, Gilbert et al. [7] presented an oddball stimulus in the presence of a series of

identically colored background items. Even though the size of the physical difference between the oddball and the background items was held constant, participants were quicker to identify the oddball when the background items belonged to a different color category from the oddball than when oddball and targets were different exemplars of the same color category. Brown et al. [8] have failed to replicate these findings and claim that color discrimination on this task is determined by perceptual rather than categorical differences. Nevertheless, similar results to those reported by Gilbert et al. [7] have been reported in several studies [9–11]. For example, Franklin et al. [10] adapted the task by presenting participants with a colored background and asking them to indicate as quickly as possible which side of fixation a differently colored patch suddenly appeared against this background. Response times were significantly faster when background color and patch came from different categories than when the patch and background were different exemplars of the same category.

In some languages such as Russian, Greek, and Korean, additional color categories exist that are not present in English. For example, *siniy* (dark blue) and *goluboy* (light blue) are distinct basic color terms for speakers of Russian, and *yeondu* (yellow-green) and *chorok* (green) are distinct basic color terms for speakers of Korean. An important question is whether CP can be observed in visual search tasks at boundaries between color categories of this kind. This issue has recently been investigated with speakers of Russian [11] and with speakers of Korean [9]. Russian participants showed CP at the boundary between *siniy* and *goluboy* in the visual search task [11]. Conversely, English speakers, who would call all of these stimuli “blue,” did not show the same cross-category advantage at the *siniy*-*goluboy* boundary. Korean participants showed CP at the boundary between *yeondu* and *chorok*, but no evidence of CP was shown by native English speakers at this boundary [9]. These findings reinforce the claim that color categories are determined by the color vocabulary that speakers acquire during language development.

Two sets of additional findings have provided information about the precise point at which categorical codes influence color categorization in this experimental paradigm. First, CP was not observed in perceptual tasks when participants carried out a concurrent verbal interference task [7, 11]; under verbal suppression, all equally spaced separations of color were equally easy to discriminate regardless of the presence of a color boundary. Second, Gilbert et al. [7] and Roberson et al. [9] reported that the CP effect was only found for colors that were presented in the right visual field, which is presumed to preferentially access language-processing areas in the left hemisphere. No difference between within-category and between-category pairings of targets and distractors was observed for colors presented in the left visual field, which gains preferential access to the nonverbal right hemisphere. Gilbert et al. [7] also showed that CP was found only when stimuli were presented to the left hemisphere of a patient in whom the corpus callosum (the structure that connects the two hemispheres of the brain) had been surgically severed.

The speed with which categorical information is accessed in these visual processing tasks suggests that there is rapid automatic retrieval of a categorical code when a colored stimulus is presented. Nevertheless, the results from these tasks can be readily explained in terms of Hanley and Roberson's [3] dual code account of categorical perception. Assume that decisions about whether a target and a background item are the same color are taken on the basis of either a right hemisphere perceptual code or a left-hemisphere categorical code and that when the two codes conflict, accuracy and speed will be reduced. Automatic activation of color category names should therefore slow judgments about whether, for example, two different shades of blue are different. This is because the categorical information that they are the same is in conflict with the perceptual information that they are different. Decisions for items from different categories (e.g., blue and green) will be quicker and more accurate because both the categorical and perceptual codes provide evidence that is consistent. When the left-hemisphere

language system is suppressed by verbal interference, or is not accessed because information is presented directly to the right hemisphere, the categorical code is not generated, and there is never any source of conflict with the perceptual code. Hence, without the language system, there is no advantage for comparisons that fall across category boundaries. Such an account can also explain why an unpredicted improvement in within-category performance has been observed when verbal interference abolished CP for color [7, 11]. Within-category performance cannot be delayed by a mismatch between the categorical and perceptual codes if a categorical code is not generated.

Categorical Perception and Event Related Potentials (ERPs)

Categorical effects for color have also been demonstrated in a paradigm that measures event-related potentials [12]. Participants were asked to look at a display containing a series of colored patches and respond as soon as they saw a cartoon character that appeared infrequently among the patches. The key manipulation was that the patches changed color; sometimes the color category changed, and sometimes the shade changed but the color category remained the same. These color changes took place at a different time from the appearance of the cartoon figures. Color processing was incidental because participants were never asked to pay any attention to color when making their responses to the appearance of the cartoon figures. The latency of the response in the ERP signal to cross- versus within-category color changes was investigated. Crucially, significantly earlier ERP peak latencies were observed when the color change involved a different color category (195 ms) than when the change involved a different exemplar from the same color category (214 ms). Such findings make it clear that information about color categories is available within the brain at a very early stage of visual processing.

Thierry et al. [13] employed a similar task to native Greek speakers. The Greek language

makes a categorical distinction between light blue (*ghalazio*) and dark blue (*ble*). Thierry et al. investigated the strength of the ERP response to unexpected changes from *ble* to *ghalazio* and reported a difference in the amplitude of the ERP. They referred to the response to such changes as a visual mismatch negativity effect (VMMN). The Greek speakers' ERP responses were significantly weaker when there was an unexpected change from light to dark green that did not cross a category boundary for either Greek or English speakers. Conversely, speakers of English did not show a stronger VMMN response to changes from *ble* to *ghalazio* than to changes from light to dark green. Athanasopoulos et al. [14] subsequently found that the signal strength of the VMMN response shown by their Greek participants when *ble* changed to *ghalazio* was negatively correlated with their length of stay in the UK. Once again, these findings emphasize the close link between visual processing of color categories and participants' linguistic experience.

Mo, Xu, Kay, and Tan [15] investigated the strength of the VMMN effect when the color change occurred in only the right (RVF) or in only the left visual field (LVF). They reported significant VMMN effects in both hemifields. Like Thierry et al. [13], Mo et al. found a stronger VMMN when a color change crossed a category boundary (in this case the boundary between blue and green) than when there was a change between two equally distinct exemplars of the same color category (e.g., different shades of green). Crucially, however, this categorical effect was observed only when the color change occurred in RVF; in LVF the strength of the VMMN was equivalent for within- and between-category changes.

In the ERP data, differential responses to cross-versus within-category pairs of stimuli appear between 100 and 300 ms after stimulus presentation. Such rapid processing of categorical information has led to suggestions that CP in perceptual tasks of this kind may reflect activity at a site in visual cortex rather than in language areas [12]. These findings are clearly consistent with the claim that CP effects for color can

sometimes be genuinely perceptual and may arise from increased perceptual sensitivity at color category boundaries. It is, however, difficult to reconcile this claim with the finding that even the early and automatic categorical color effects that are detected by ERPs are lateralized to the left cerebral hemisphere [15]. Furthermore, the published studies that employ these ERP tasks have, to date, only reported cross-category versus within-category contrasts. Exactly the same perceptual discriminations are required regardless of whether the color change on within-category trials is from a good category member to a peripheral category member or vice versa. It remains to be seen whether the ERP signal is equally weak in both these situations relative to the ERP signal generated by a color change that crosses a category boundary. It is therefore unclear whether the asymmetry of within-category performance found consistently to cause CP in 2-AFC tasks [2] generalizes to visual mismatch negativity in ERP signals. If color CP effects in these visual processing tasks are also the result of a category contrast effect, it would be difficult to see how an explanation of the ERP data that is based on perceptual sensitivity could reasonably be maintained.

Cross-References

- [Berlin and Kay Theory](#)
- [Infant Color Categories](#)

References

1. Roberson, D., Davies, I., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
2. Hanley, J.R., Roberson, D.: Categorical perception effects reflect differences in typicality on within-category trials. *Psychon. Bull. Rev.* **18**, 355–363 (2011)
3. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)

4. Harnad, S.: Psychophysical and cognitive aspects of categorical perception: a critical overview. In: Harnad, S. (ed.) *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
5. Hampton, J.A., Estes, Z., Simmons, C.L.: Comparison and contrast in perceptual categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1459–1476 (2005)
6. Stewart, N., Brown, G.D.A., Chater, N.: Sequence effects in categorization of simple perceptual stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 3–11 (2002)
7. Gilbert, A.L., Regier, T., Kay, P., Ivry, R.B.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 489–494 (2006)
8. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes color perception is not categorical. *J. Vis.* **11**, 1–21 (2011)
9. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of color in the left and right hemisphere is verbally mediated: evidence from Korean. *Cognition*. **107**, 752–762 (2008)
10. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T.: Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3221–3225 (2008)
11. Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Boroditsky, L.: Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7780–7785 (2007)
12. Fonteneau, E., Davidoff, J.: Neural correlates of color category perception. *Neuroreport*. **18**(13), 1323–1327 (2007)
13. Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., Kuipers, J.R.: Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4567–4570 (2009)
14. Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., Thierry, G.: Perceptual shift in bilingualism: brain potentials reveal plasticity in pre-attentive color perception. *Cognition*. **116**, 437–443 (2010)
15. Mo, L., Xu, G., Kay, P., Tan, L.-H.: Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14026–14030 (2011)

Color Categories

► Color Categories and Vagueness

Color Categories and Vagueness

Igor Douven

SND/CNRS/Sorbonne University, Paris, France

Synonyms

[Color categories](#); [Color terms](#); [Conceptual spaces](#); [Logic](#); [Vagueness](#)

Definition

Natural language is full of vagueness. Some people are definitely tall, others are definitely *not* tall, but there are also people who possess the property of tallness only to some intermediate degree. While colors are paradigmatic examples of properties that objects may possess to some intermediate degree, most of the work on color categorization has almost completely ignored the issue of vagueness. The issue has been addressed in more recent research, however, and this research has also contributed to our general understanding of vagueness.

The Metaphysics of Color Categories, Their Prototypes, and Their Boundaries

We carve up reality into different categories for both intellectual and practical reasons: being able to differentiate among the many objects we encounter facilitates learning, memorization, prediction, and action. There is an ongoing debate about the metaphysical status of those categories. Do they reflect fundamental building blocks of the world – at least ideally, leaving open the possibility that our *current* categories fall short of meeting that ideal – or are they merely conventional, and would other ways of carving up reality have served us just as well? In the former case, categories would be universal, at least under ideal circumstances; in the latter case, one would expect them to be relative to culture.

In philosophy and beyond, there has been much a priori speculation about the universalism–relativism issue. Twentieth-century color research, in particular the work on color naming initiated by Berlin and Kay [1], provided important *empirical* input to the debate (→ “► [Berlin and Kay Theory](#)”; → “► [Color Categorization: Patterns and Mechanisms of Evolution](#)”). And while there was evidence in Berlin and Kay’s data for there being conventional aspects to color categorization (→ “► [Cross-cultural Communication](#)”), the preponderance of evidence was in favor of a universalist conception of color categorization, leading Berlin and Kay to posit the existence of so-called basic color categories (BCCs). Thereby, their findings were also at least partial evidence for universalism vis-à-vis categorization more generally.

Related to the debate about the metaphysical status of categories has been a debate about prototypes, which are generally understood as being the categories’ most typical representants. Might these be best thought of as being conventionally determined – perhaps depending on which instances we have most frequently, or most recently, encountered – or are they also part of the blueprint of reality [2, 3]? Note that this issue is to some extent orthogonal to the question of the status of categories: even if the categories are universal, the choice of which instances best represent those categories might be entirely up to us, so that different cultures with a shared system of categories might still identify different instances as being prototypical for those categories. Here, too, Berlin and Kay’s work provided empirical support for a universalist position, at least in the domain of color.

More recently, the empirical results have been buttressed by various results coming out of a computational modeling approach to categorization, especially color categorization (→ “► [Color Category Learning in Naming-Game Simulations](#)”). The most notable results to date were obtained in important work by Regier and colleagues [4], who – following an idea from Jameson and D’Andrade [5] – implemented a number of general information-theoretic principles in an optimization algorithm. Specifically, Regier and

colleagues placed the 330 color chips that Berlin, Kay, and their collaborators had used (→ “► [World Color Survey](#)”) in CIELAB space (→ “► [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)”; → “► [Psychological Color Space and Color Terms](#)”) and let a computer divide those chips into 3, 4, 5, and 6 categories, subject to the constraint that the resulting system should maximize similarity of the colors within each category while also minimizing similarity of the colors across the different categories, where similarity was defined in terms of closeness in CIELAB space. Regier and colleagues were thus able to recover computationally many of the color categorizations that Berlin, Kay, and their collaborators had found in their field work.

However important the aforementioned research is, it can easily leave one with a wrong idea about the *boundaries* of color categories. In standard ways of presenting results concerning color categories, these categories almost invariably have *sharp* boundaries, as if there were regions in color space such that, for instance, any shade represented by a point in that region is definitely orange while any shade represented by a point outside that region is definitely not orange. That color categories are so sharply delineated from one another may be a convenient assumption for a variety of purposes, but there can be no question that it is false. Most drawing software allows one to pick a color and then gradually change its coordinates (typically, its RGB coordinates). Assuming we have access to such software, we can, for example, start with a clear shade of red, then adjust the color coordinates little by little such that the shades we are seeing gradually become more orangish, until we arrive at a shade that we would designate as being clearly orange. The shades we saw in the process, which can be thought of describing a path through color space, will certainly not all have been clearly identifiable as red or clearly identifiable as orange. Many will have been somehow in-between those colors; their color will have been *vague*.

Though it is evident, at least upon reflection, that color categories have vague rather than sharp boundaries, and that hence predicates like “red”

and “blue” sometimes apply only to some intermediate degree, the study of the issue of vagueness in relation to color is badly neglected. One reason for this may be that the issue of vagueness itself is a can of worms, with linguists, logicians, and philosophers having proposed a number of ways of dealing with vagueness, none of which has been able to garner widespread support.

Language, Logic, and Vagueness

Modern logic was developed to facilitate mathematical reasoning. As such, it has been immensely successful. But in a number of respects, the language of mathematics is special. Most notably, there are no vague mathematical expressions. That is a problem if we seek to extend the scope of logic to natural language. In standard logic, there is the law of excluded middle, which implies that for any predicate P and any constant c , $P(c)$ is either true or false. But for many ordinary language predicates, that law appears to fail. As mentioned, for instance, there are people we deem neither definitely tall nor definitely not tall, and similarly for many color shades, which we do not deem to fall squarely into any one color category. In the literature on vagueness, such cases are referred to as *borderline cases*.

Logicians have responded to this problem in two ways: (i) deny the existence of vagueness, and (ii) revise logic. The first response [6] asserts that, for any seemingly vague predicate, there is actually a precise cut-off point, delineating the items to which the predicate applies from those to which it does not apply. In this view, vagueness is not really a metaphysical problem but only an epistemic problem. It is an expression of our ignorance, stemming from our limited discriminatory capacities, and not a feature of the world.

The second response comes in a variety of flavors. Some have proposed to admit a third truth value, which is then to indicate something like indeterminacy. Others have wanted to go beyond this, by admitting *many* intermediate truth values, possibly even *infinitely* many. (See the papers in [7] for useful discussion of the various options.)

While none of these proposals could be said to be popular, there is a proposal for revising semantics – roughly understood as a theory of assigning truth values to sentences, which is an important part of logic – that has recently received some empirical support. Kamp and Partee [8] consider a language that contains both vague and nonvague predicates and show how one can obtain a unified semantics for that language. One of the cornerstones of their proposal is the notion of a *completion*, which is a way of resolving vagueness by arbitrarily choosing, for any vague predicate P , a cut-off point such that some of the borderline P cases are counted as being definitely P while the remaining borderline cases are counted as being definitely *not* P . While the cut-off point is arbitrary, it should still be the case that if a borderline P case a is counted as being definitely P , then any borderline case b that is at least as P (at least as tall, green, or whatever P stands for) as a should also be counted as being definitely P .

The second cornerstone of Kamp and Partee’s proposal is the notion of *graded membership*, which they define in terms of completions. Suppose a simple case in which there are only finitely many completions, that is, finitely many ways of splitting up the borderline cases such that the aforementioned constraint (if a is definitely P then anything at least as P as a is definitely P as well) is met. Then for any borderline P case a , we can take the proportion of completions that group a with P as measuring the degree to which a is P . If there are infinitely many completions, then the idea is the same, although the implementation requires the help of measure theory.

Kamp and Partee’s proposal had some loose ends. It turned out, however, that the remaining problems could be easily solved by embedding their proposal in a conceptual spaces framework [9]. Doing so has the additional advantage of making the proposal testable. Color space being one of the most readily available conceptual spaces (e.g., CIELAB and CIELUV space are accessible via a number of widely used software packages; → “► [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)”), it is unsurprising that some of the testing has taken place in the color domain.

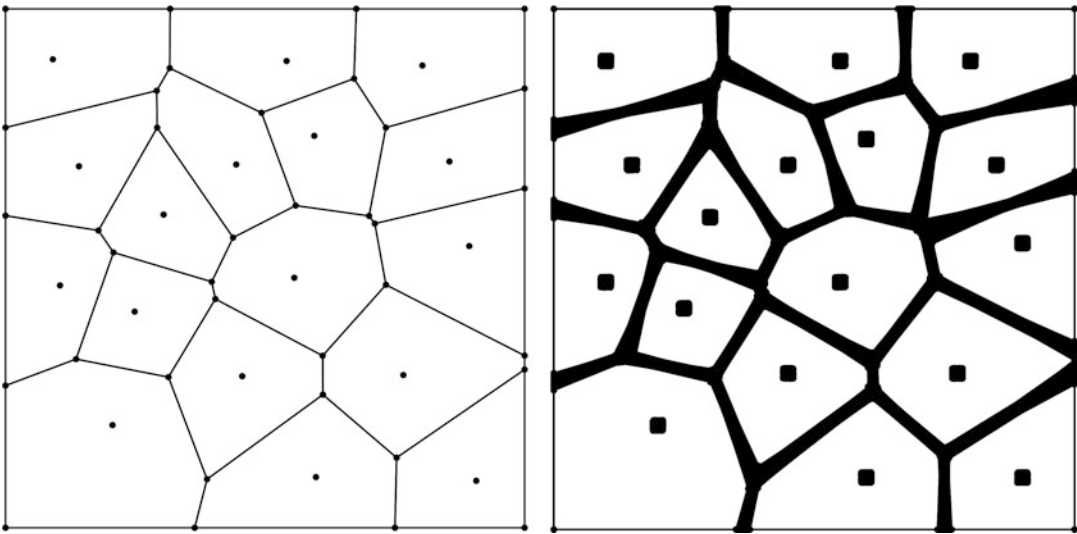
Predicting the Vagueness of Color Shades

Conceptual spaces are built on top of similarity spaces, which are one- or multidimensional metric spaces in which distances measure dissimilarities [10]. The further apart two points are in a similarity space, the more the items mapped onto those points are dissimilar in the respect represented by the space. Using as an example some color similarity space (e.g., CIELAB or CIELUV space), the closer two points are in this space, the more similar in color are items mapped onto these points [11].

To obtain a conceptual space from a similarity space, we add a structure to the latter that carves it up into regions, where these regions are then taken to represent concepts, generally understood as the mental correlates of linguistic expressions. For example, a conceptual color space is a carving up of CIELAB or CIELUV space, or possibly some other color similarity space, into a number of regions, with the various regions representing color concepts. In this picture, the concept of redness can be thought of as a region in CIELAB or CIELUV space, and as can for the concepts of blueness, greenness, and so on.

A popular way of arriving at such a carving-up uses the mathematical technique of *Voronoi tessellations*. A Voronoi tessellation starts with a set of *generator points* in a similarity space, and then divides that space into as many cells as there are such points, following the principle that, for any cell, all points lying in the cell should be at least as close to the generator point associated with that cell as they are to any of the other generator points. A natural choice of generator points in the context of conceptual spaces has been prototypes. For instance, to arrive at a conceptual color space, we might determine the locations of the BCC prototypes in, say, CIELAB space and then let these serve as generator points for a Voronoi tessellation of that space. See the left panel of Fig. 1 for an illustration.

The tessellations resulting from this kind of operation are known to have various attractive features from a cognitive point of view [10]. Depending on the concepts one wants to represent, however, a drawback can be that this type of tessellation makes the boundaries between concepts *sharp*. For color concepts, the drawback is particularly evident, for a reason already mentioned: if we traverse CIELAB or CIELUV space,



Color Categories and Vagueness, Fig. 1 Left: A Voronoi tessellation of a two-dimensional space, the points representing the prototypes and the lines showing the borders between the categories. Right: A collated

Voronoi tessellation, equipped with prototypical regions instead of prototypical points, and having thick boundary regions instead of sharp borders between categories

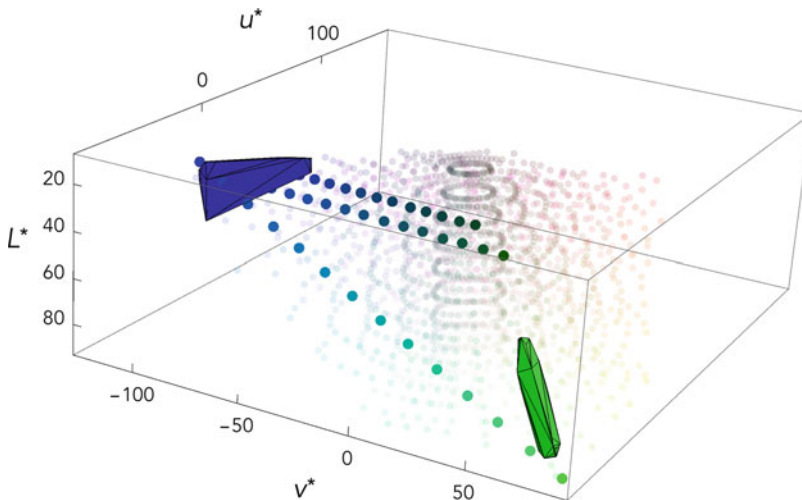
we do not find the abrupt transitions between, for instance, red and orange that are suggested by a Voronoi tessellation of those spaces obtained in the way just described. Instead, we find gradual transitions; for instance, traveling from clearly red to clearly orange, we go through various stages of borderline red-orange.

To solve this problem, Decock and Douven [9] draw on *collated* Voronoi tessellations, which pack together bundles of (single) Voronoi tessellations, melding them into one tessellation [11]. This leads to concepts being separated from one another by boundary *regions* rather than boundary *lines*. Their proposal is to first do away with the fiction that concepts always have a singular prototype (e.g., there are many shades of red that people deem typically red, as is also attested by empirical research [1, 3]) and rather recognize that, at least for many concepts, we can discern *regions* in a conceptual space where their prototypical items are represented. Then a collated Voronoi tessellation is constructed by conjoining all ordinary Voronoi tessellations that are generated by a set of points consisting of one from each prototypical region. These have boundary regions which meet a number of important desiderata

[12]. See the right panel of Fig. 1 for an illustration.

The final step in Decock and Douven's proposal is to connect the idea of conceptual spaces equipped with a collated Voronoi tessellation with Kamp and Partee's semantics for languages with vague terms. The main bridge between the two lies in the observation that the single Voronoi tessellations that make up a collated tessellation can be regarded as the correlates in the conceptual spaces framework of Kamp and Partee's completions: each such single tessellation is a way of grouping items in the boundary region of a concept with either the clear instances of that concept or its clear non-instances, and – because of the metric structure of the space – it does so in a way that satisfies Kamp and Partee's constraint on completions.

An immediate empirical consequence of this is that degrees of membership of any item representable in a space can be calculated from the structure of the space together with the locations in the space of the prototypical regions of the concepts it represents. For instance, given that the structure of the main color similarity spaces is known, knowledge of where in those space the regions



Color Categories and Vagueness, Fig. 2 CIELUV space with the prototypical regions for blue and green, and with three series of shades all going from clearly blue to clearly green. The highlighted points are shades for which researchers ([13]) calculated the degree of blueness and degree of greenness, which they then compared with

human judgments. (To give an impression of the global shape of CIELUV space, the figures also show thinly the set of 1625 Munsell chips available from the website of the Program of Color Science, Munsell Color Science Laboratory, Rochester Institute of Technology)

representing the prototypical red shades, the prototypical orange shades, and so on, are located allows one to calculate, for any point in the space, the degree to which the shade corresponding to that point is red, orange, blue, green, and so on.

There is recent empirical evidence that bears out the predictions flowing from the above proposal. Specifically, researchers [13] gathered information about the locations of the prototypical regions for blue and green in CIELUV space; see Fig. 2. They then calculated, for various shades in the blue-green region of that space, also shown in the figure, the degree to which they were green and the degree to which they were blue, according to Kamp and Partee's semantics as embedded in the conceptual spaces framework. The predictions turned out to be highly accurate, thereby buttressing that semantics and making a strong empirical case for a particular account of vagueness.

Cross-References

- [Berlin and Kay Theory](#)
- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Color Category Learning in Naming-Game Simulations](#)
- [Cross-cultural Communication](#)
- [Psychological Color Space and Color Terms](#)
- [World Color Survey](#)

References

1. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley/Los Angeles (1969/1991)
2. Rosch, E.: Natural categories. *Cogn. Psychol.* **4**, 328–350 (1973)
3. Douven, I.: Putting prototypes in place. *Cognition*. **193**, 104007 (2019). <https://doi.org/10.1016/j.cognition.2019.104007>

4. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1436–1441 (2007)
5. Jameson, K., D'Andrade, R.: It's not really red, green, yellow, blue: an inquiry into perceptual color space. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 295–319. Cambridge University Press, Cambridge (1997)
6. Williamson, T.: *Vagueness*. Routledge, London (1994)
7. Dietz, R., Moruzzi, S.: *Cuts and Clouds: Vagueness, Its Nature and Its Logic*. Oxford University Press, Oxford (2010)
8. Kamp, H., Partee, B.: Prototype theory and compositionality. *Cognition*. **57**, 129–191 (1995)
9. Decock, L., Douven, I.: What is graded membership? *Nous*. **48**, 653–682 (2014)
10. Gärdenfors, P.: *Conceptual Spaces*. MIT Press, Cambridge, MA (2000)
11. Fairchild, M.D.: *Color Appearance Models*. Wiley, Hoboken (2013)
12. Douven, I., Decock, L., Dietz, R., Egré, P.: Vagueness: a conceptual spaces approach. *J. Philos. Log.* **42**, 137–160 (2013)
13. Douven, I., Wenmackers, S., Jraissati, Y., Decock, L.: Measuring graded membership: the case of color. *Cogn. Sci.* **41**, 686–722 (2017)

Color Categorization

- [Chinese Color Language](#)
- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Color Constancy and Categories](#)
- [Color Name Applications in Computer Vision](#)

Color Categorization and Naming in Inherited Color Vision Deficiencies

Valerie Bonnardel
Psychology, University of Winchester,
Winchester, UK

Synonyms

[Dichromat color naming](#)

Definition

Color categorization and naming behaviors of human observers that experience forms of color vision deficiency called “dichromacy.” Such deficiencies are sex linked and predominantly affect males and are due to errors in photopigment expression or functioning or to the failure to inherit the genetic precursors for the expression of a normal set of retinal photopigments.

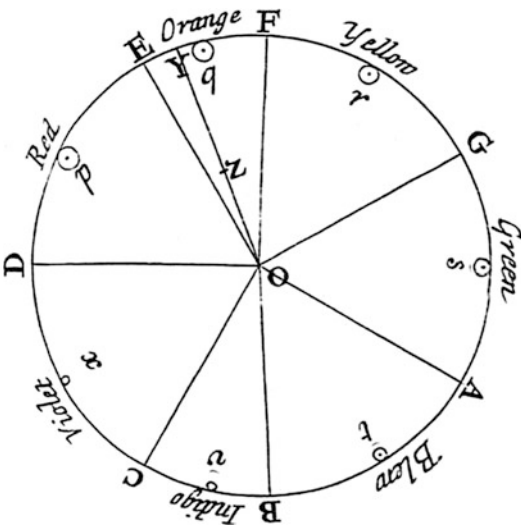
Color Naming and Categorization in Color Deficients

Human color perception is categorical. Light diffracted through a prism provides a visible spectrum spanning continuously from 400 to 700 nm in the wavelength interval that readily appears to the human eye as a smooth juxtaposition of colored bands of different width. By adding to the visible spectrum a mixture of spectral lights, Isaac Newton described a color circle with its different sectors as *violet*, *indigo*, *blue*, *green*, *yellow*, *orange*, and *red* (Fig. 1) [1]. For the great majority of the Caucasian population (96%), Newton’s

seven-color categories (although not unanimously agreed upon) are a fair account of human color experience. However, for a substantial percentage of men who are color deficient (8%), this description is inadequate. Compared to individuals with normal color vision, color deficient have reduced color discrimination abilities and live in a linguistic environment with a larger color vocabulary than the gamut of color percepts they probably experience. Yet, in everyday life, color deficient’s use of color names permits intelligible use of color in conversation, and their deficiency is largely unnoticeable by color normal interlocutors. As such, color deficiencies offer a useful paradigm for the study of the complex relationships between idiosyncratic perceptual experience and its linguistic expression.

Color Vision Deficiencies

Only inherited forms of color vision deficiency associated with changes in the genes on the X-chromosomes will be considered here. Normal color vision is trichromatic and results from light absorption by three different types of photopigments located in the retinal cell receptors called “cones.” Short-wavelength (SW), medium-wavelength (MW), and long-wavelength (LW) cone photopigments absorb light maximally in short (440 nm), medium (540 nm), and long (560 nm) wavelengths. A genetic polymorphism affecting X-linked genes encoding for MW and LW cone photopigments induces small variations in their absorption peak, leading to subtle variations among normal trichromats; larger variations produce deficiencies. In the extreme case, a gene deletion occurs, and as a result, color vision will lose one dimension and will be dichromatic. A more moderate condition arises from gene alterations that generate hybrid photopigments. In this case, color vision is referred to as anomalous trichromatic. When LW photopigment is affected, deficiencies are of the protan type, with about 1% of dichromats (of “protanope” type) and 1% of anomalous trichromats or “protanomalous.” When the MW photopigment is concerned, deficiencies are of deutan type and account for frequencies of 1.4% “deuteranope” and 4.6% of “deuteranomaly.”



Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 1 Color circle from Newton’s Opticks [1]. (Source: <http://posner.library.cmu.edu/Posner/books/>)

In practice, color deficient people confuse colors that are easily discriminable to normal trichromats. To gain intuition on color deficient's perceptual experience, two clarifications are useful to bear in mind. The first consists in defining the term "color" which can be polysemous. Color can be defined as corresponding to the appearance of objects and lights resulting from the combination of three perceptual attributes associated to three physical variables, that is, "hue" (associated to dominant or complementary wavelength), saturation (associated to purity), and brightness (associated to luminance). Color deficient people primarily confuse hues and retain the ability to discriminate between different levels of saturation and brightness. Secondly, photoreceptor absorption process is not useful when considering the perceptual implications of their dysfunction. This is better accounted for by considering further stages of color processing which involve synergic and antagonist neural linkages between photoreceptors to form an achromatic (white vs. black) [MW+LW] and two-chromatic channels: "red" versus "green" [LW-MW] and "yellow" versus "blue" [(LW+MW) – SW]. Dichromats who are missing either LW or MW photoreceptors are assumed to retain a single color-opponent channel: [MW-SW] or [LW-SW] (protanope or deuteranope, respectively). Anomalous trichromats possess a residual yet functional "red-green" [LW-MW] channel. Considering the diversity of

anomalous trichromat types, illustrations of color deficiencies will be limited to dichromatic vision.

Algorithms based on the reduction assumption permit dichromatic color gamut simulations as seen by individuals with normal color vision [2]. The simulated color circles for deuteranope and protanope are filled with blue, yellow, and shades of those colors and a neutral or achromatic zone (Fig. 2). Although the qualities of another person's sensations can never be fully known, such simulations provide insights on magnitudes of the difference that might exist between dichromatic and trichromatic color experience.

Despite substantial differences in color experience, the condition of color deficiency had not been reported until the end of the eighteenth century. John Dalton (posthumously diagnosed in 1995 as a deuteranope by means of molecular genetic techniques [3]) provided the most comprehensive description of color deficiency based on his own experience. It is at age of 28, when noticing that a geranium changed its color from sky blue in daylight to yellowish in candlelight, that Dalton suspected his color vision to be different from that of others. From systematic observations of the solar spectrum, Dalton established that if people distinguished six color categories, namely, *red*, *orange*, *yellow*, *green*, *blue*, and *purple* (further divided in *blue* and *indigo* to fit with Newton's nomenclature), he was able to see only two or at most three distinctions: *yellow* and *blue*



Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 2 Color circle obtained from sRGB color display (center) and its two simulations as

supposedly seen by a deuteranope (left) and protanope (right). Simulations were obtained with Vischeck (www.vischeck.com)

or *yellow*, *blue*, and *purple*. This description is in fair agreement with today simulations presented in Fig. 2.

Dichromats Color Naming and Categorization Abilities

In everyday life, colors are usually object colors, and as young children, dichromats can learn to associate color names with objects of pre-determined color such as green grass or red cherries, and it is not surprising that dichromats use “red” and “green” to refer to colors that seem so distinctive to others. However, “red” and “green” are also used accurately most of the time for non-predetermined object colors such as clothes or furnishings, and asking color names to color deficient gives little indication of a color perception deficiency. It is in laboratory situation that one can measure the discrepancy between perception and naming in dichromats. Jameson and Hurvich [4] in an article entitled “Dichromatic color language: “reds” and “greens” don’t look alike but their colors do” [4] reported an experiment using arrangement color tests (i.e., Farnsworth D-15, with 15 colors), where colors printed on small caps should be ordered by perceived similarity. In this type of tests, dichromats make systematic and identifiable color alternations from which color diagnosis is based. For instance, dichromats put cap #2, considered by normal trichromats as “blue-green,” next to cap #13, a “violet-blue” cap. When subsequently, dichromatic observers were asked to name the same stimuli, the authors noted several interesting observations. Firstly, and despite confusing their colors, all dichromats used “red” and “green” words. Secondly, a dichromat could describe a color cap as “red-green.” This designation never occurs in normal trichromat observer as, due to their antagonist nature, these percepts are mutually exclusive. Thirdly, a large interindividual variation was reported among dichromats. For the less keen observers, “red”-“green” were used indifferently to refer to colors seen as red or

green by normal trichromats. Yet, some dichromats were able to produce a naming pattern indistinguishable from that of a normal trichromat, thus revealing a gap between perceived similarity and naming. For instance, if cap #2 was placed next to cap #13 in the arrangement test, in agreement with normal trichromat, cap #2 was named “blue-light” and cap #13 “reddish blue.” In this experiment, correct naming was explained by subtle luminance differences between red and green caps that could have been used by dichromats. To the careful dichromatic observer, the simple rule “darker, then red” provides a better-than-chance correct naming.

This rule would be more difficult to apply when colors are shown in isolation. This setup was used in an experiment with 140 Munsell color samples varying in hue, saturation, and brightness intended to probe naming and categorization. In the naming task, participants had to choose one of the eight chromatic basic color terms (BCT, namely, *red*, *green*, *blue*, *yellow*, *orange*, *pink*, *purple*, and *brown*) to describe each sample presented in isolation. Categories thus obtained for the two dichromats tested were very similar to normal trichromatic prototypical naming categories, with 66% and 72% of the samples described with the same BCT as used by normal trichromats. In the categorization task, color samples were sorted in eight categories based on their perceptual similarity with, this time, the overall sample collection displayed on a large table, inciting comparison strategy. Despite the possibility of samples’ visual comparison, local category inversions corresponding to cases where stimuli were assigned to groups in a way that is contrary to the structure of the nearest neighbor were observed in the categorization task, while absent in the naming task. Moreover, for the same number of categories, color naming produced a more consensual categorization pattern compared to perceptual similarity. Indeed, color names that define category boundaries in terms of hue (i.e., red, green, blue, yellow, orange, and purple), saturation (i.e., pink), and brightness (i.e., brown) further constrain the elicited

categories by, for instance, excluding categories such as turquoise or pastel that would be legitimate based on similarity criteria. This added constraint in normal trichromats increased the consensus between participants from 72% (categorization task) to 82% (naming task) and similarly from 66% to 76% in color deficient [5]. Language appears to operate normalization effect on categories obtained by naming in color deficient, yet it does not seem to be readily available to fine-tune color categories when these are based on perceptual similarities.

Explanations for Dichromat Color Naming and Categorization Abilities

Aside from subtle luminance differences or lightness cues, dichromat color naming and categorization performances have been explained by the contribution of other visual signals. In particular, rods distributed in peripheral retina and absent from the fovea are mediating low-light intensity vision and can act as a third photoreceptor. Under moderate light conditions, for surfaces larger than 4° of visual angle or presented in periphery, dichromats reveal trichromatic color matches comparable to normal trichromats' small-field (2°) color match, that is, for these larger field sizes, they need two primary lights (red and green) to match a spectral yellow light. Under these conditions, dichromat large-field trichromacy is explained by rod-cone interaction providing a residual red-green sensitivity [6]. In the naming of 424 OSA Uniform Scales samples presented in isolation, dichromats produced trichromat-like responses in conditions where rod-cone interactions were possible. When rod contribution was excluded by bleaching the photopigment by light preexposure, trichromat-like naming was no longer possible in the red-green dimension [7].

Dichromats still perform better than expected on perceptual tasks in conditions where luminance and rod-cone interaction are ruled out. For instance, in hue-scaling experiment, stimuli were monochromatic lights presented in a 2° field. Observers' task consisted in describing each

stimulus by a given proportion of "yellow," "blue," "red," and "green." In this experiment, a molecular genetic analysis had confirmed that dichromat participants had only one functional X-chromosome-linked photopigment opsin gene (either L or M), excluding the possible contribution of a residual red-green color vision. Dichromats reported the presence of a red component in the 420–450 nm interval and explicitly referred to blue and red mixture. These data suggest a dichromatic color space structure is richer than that illustrated by simulations based on linear model and reduction assumption as presented in Fig. 2. A model including a nonlinear transfer function of signals from S and L or M cones can account for a richer dichromatic perceptual space with enhanced naming and categorization abilities [8].

Non-chromatic visual cues, rod-cone interaction, or nonlinear transfer function could all provide supplementary visual information to dichromats to resolve ambiguities in an otherwise impoverished chromatic environment. However, color naming shows a more trichromat-like pattern than similarity or perceptual categorization tasks can elicit. The comparison between colors and color-naming structures has been addressed in similarity tasks using color cards and their corresponding color names printed on cards. Similarity judgments were requested for 36 pairs of stimuli. Apart from a violet-blue inversion, the multidimensional scaling analysis of dichromat color name data provided a Newton color circle in a two-dimensional model with yellow-blue and red-green axis. For color similarities, the color circle was distorted along the red-green axis, bringing these two colors next to each other [9].

Dichromats have correctly learned the relational structure between color names as established by normal trichromats. Yet, as noted by Jameson and Hurvich [4], a dichromat's correct conceptual representation of color is not used to optimize performance on perceptual tasks. This finding indicates that the existence of an isomorphism between percept and concept structures is not compulsory; each type of representation coexists with no apparent conflict and can be used independently to fulfill specific task demands.

Cross-References

- [Color Categorical Perception](#)
- [Cone Fundamentals](#)
- [Environmental Influences on Color Vision](#)
- [World Color Survey](#)

References

1. Newton, I.: *Opticks: or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. Smith & Walford, London (1704)
2. Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J.D.: What do color-blind people see? *Nature*. **376**, 127–128 (1995)
3. Hunt, D.M., Dulai, K.S., Bowmaker, J.K., Mollon, J.D.: The chemistry of John Dalton's color blindness. *Science*. **267**, 933–1064 (1995)
4. Jameson, D., Hurvich, L.M.: Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sens. Process.* **2**, 146–155 (1978)
5. Bonnardel, V.: Naming and categorization in inherited color vision deficiencies. *Vis. Neurosci.* **23**, 637–643 (2006)
6. Smith, V.C., Pokorny, J.: Large-field trichromacy in protanopes and deuteranopes. *J. Opt. Soc. Am.* **67**, 213–220 (1977)
7. Montag, E.D., Boynton, R.M.: Rod influence dichromatic surface color perception. *Vis. Res.* **27**, 2153–2152 (1987)
8. Wachtler, T., Dohrmann, U., Hertel, R.: Modeling color percepts of dichromats. *Vis. Res.* **44**, 2843–2855 (2004)
9. Shepard, R.N., Cooper, L.A.: Representation of colors in the blind, color-blind, and normally sighted. *Psychol. Sci.* **3**, 97–110 (1992)

Color Categorization: Patterns and Mechanisms of Evolution

Galina V. Paramei
Department of Psychology, Liverpool Hope
University, Liverpool, UK

Synonyms

[Basic color categories](#); [Basic color terms](#); [Berlin-Kay hypothesis](#); [Color categorization](#); [Culturally specific basic color terms](#); [Universalism-relativism debate](#)

Definition

Color categories are mental constructs that partition the gamut of all colors, perceived by an individual, into a number of discrete entities, each labeled with a color name [1, 2] (► [“Color Categorical Perception”](#)). In anthropologic and psycholinguistic studies, a color category/name is defined in terms of its objective referents, or denotata, i.e., a range of color samples it encompasses. The referents are either standardized color chips, such as those constituting Munsell charts (► [“World Color Survey”](#)), or color stimuli whose hue, brightness/lightness, and saturation are specified in photometric terms (e.g., CIE x y z , CIELUV, CIELAB, etc.) [3] (► [“CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength”](#)). The volume of color categories and the boundaries that demarcate them manifest by consensus across the speakers of a language, facilitating effective communication (► [“Psychological Color Space and Color Terms”](#)). Beyond the denotative function, in linguistics, and, in particular, in linguistic pragmatics, the meaning of a color term is defined by its referential meaning, or use patterns, such as frequency of use, immediacy of response, collocations, and connotations, which reflect the term's affective values, cultural conventions, semiotics, and social-marker functions [4].

The Universalism Versus Relativism Debate

Until late 1960s, dominant was the *relativist view* (aka the Sapir-Whorf hypothesis) on color categories in individual languages. According to it, languages mark off different parts of the color continuum quite arbitrarily, implying that color names of different languages do not embrace the same denotative entities, and the only universally true generalizations to be drawn are trivial or tautological [5].

The paradigm changed to the counterhypothesis after publication of the seminal work of Berlin and Kay [6], whose *universalist view* posits panhuman basic color categories

(BCCs) (► [“Berlin and Kay Theory”](#)). For estimating BCC denotata, the authors employed a standardized array of 330 color stimuli – the “envelope” (Mercator projection) of the Munsell Color Solid, with color chips varying in hue, lightness, and being of highest saturation (see Fig. 1 in ► [“World Color Survey”](#)). In its denotative structure, a BCC is organized around its “best exemplar,” or category *focal color* (see Fig. 4 in ► [“World Color Survey”](#)), and extends to its boundary (► [“Psychological Color Space and Color Terms”](#)). Nonfocal colors in a BCC are characterized by a *graded membership*, as demonstrated by latest empirical and modeling developments [7] (► [“Color Categories and Vagueness”](#)).

According to the universalist view, BCCs recur across languages and are evolving in a partially fixed order, from a minimum of 2 to a maximum of 11 (in industrialized societies) (see Fig. 2 in ► [“World Color Survey”](#); ► [“Berlin and Kay Theory”](#)). Since 1969 the universalist view gathered abundant support from numerous studies around the world that employed the Munsell array and standardized methodology (► [“World Color Survey”](#); ► [“Color Dictionaries and Corpora”](#); ► [“Mesoamerican Color Survey Digital Archive”](#)). Based on new findings, the original model has been revised and elaborated by the authors and their collaborators, with a special regard to sequences of BCT emergence, with the latest version coined *the Universals and Evolution (UE) model* [8].

It is generally agreed that the BCCs emerge and evolve as the result of universal constraints at different levels of color processing [9]. First, the BCC structure is determined by the anatomy and physiology of the human visual system. In addition, it is stipulated by characteristics of color perception per se, articulated by Jameson and D’Andrade [10] as the *optimality hypothesis*: the BCC structure reflects a partition of color space as a way of meaningful information coding of the visible color gamut, so that in the development of the BCT inventory, color categories are added that maximize color differences between adjacent categories and minimize color differences within the new contiguous categories.

Notably, individual language speakers may have their variants of the shared structure – due to physiological differences in the visual system, and/or exposure to different chromatic environments [11], and/or perceptual and cognitive biases that modulate the general constraints (► [“Dynamics of Color Category Formation and Boundaries”](#)).

The universalist view underwent, however, criticisms from proponents of the relativist view. In recent times, based on results of multiple empirical studies, in particular, the WCS, the debate resulted in the compromise of the *weak relativity hypothesis* that reconciles the two extreme theoretical views: it accepts that both category focal colors and boundaries of BCCs may fluctuate between individual languages [12] (see Fig. 5 in ► [“World Color Survey”](#)). The *weak relativity hypothesis* acknowledges that in individual languages, color categories are also shaped by the speakers’ culture-specific, social, and pragmatic factors [2, 4, 9, 13] (► [“Cross-cultural Communication”](#)), as well as by the color diet in the population’s natural environment [14].

Computational simulations that used an extensive number of color stimuli (► [“Color Category Learning in Naming-Game Simulations”](#)) provided decisive evidence that BCCs are not arbitrary: the emergent categorization pattern is dominated by perceptual constraints of human color vision. The language and sociocultural factors are also essential in shaping the color categorization pattern and act against the pressure for uniformity, where population-specific constraints introduce categorical variation stipulated by cultural histories of speaker community.

Modern Age Phenomenon: Increasing Lexical Refinement of Color Categorization

The *weak relativity hypothesis* implies the possibility of new BCCs, which are specific to a certain language. Indeed, recent empirical studies provide accumulated evidence of additional BCCs in some languages, beyond the 11 established by Berlin and Kay [6]. Conceptual and lexical

refinement was recorded, in particular, in the blue area of color space – with “light blue” and “dark blue” BCCs existing in several Slavonic languages (Russian, Ukrainian, Belarusian), circum-Mediterranean languages (Italian, Greek, Turkish, Maltese), some Far East languages (Japanese, Thai), and in Guatemalan and Uruguayan Spanish dialects (for a review see [15]).

Analysis of the emerging BCCs in the languages with more than 11 BCTs prompts that lexical refinement of color space can unfold along one of three ways (for an overview see [15]):

1. Lexical *partition* of the color gamut (assumed in [6]), whereby a novel BCC/BCT emerges as a result of an “old” BCC undergoing fission, or splitting into two, as illustrated by the two blue categories
2. *Insertion* of a new category that straddles the boundaries of existing BCCs, with two color space areas being particularly prone to further lexical differentiation, purple-red and blue-green category boundary
3. *Emergence* of a color category in “under-named” areas of color space that become salient, so a need arises to label them

Furthermore, modern times observe intensive lexical differentiation of BCCs in economically developed cultures – by amplified entrenchment of existing non-BCTs (e.g., “turquoise,” “beige,” “salmon,” “peach,” “magenta,” “cyan,” “fuchsia,” “tan,” etc.) and emergence of new non-BCTs (► [“Color Dictionaries and Corpora”](#)). The general tendency of an increasing refinement of color-space naming is a cognitive response to cultural-economic factors in the globalized world of trade, new technological developments of colored artifacts, and media.

Along with the general factors, lexical refinement of color space is also driven by *language-specific* factors, which reflect cultural influences and visual-environmental statistics, in accord with a hypothesis of usefulness of color to a culture [2, 14, 16] (► [“Cross-cultural Communication”](#)).

Both phenomena were recently analyzed from the viewpoint of general information-theoretic

principles, with the conclusion that the evolution of color-term repertoire is primarily driven by communication constraints [17], with different researchers emphasizing communication pressure [18], communicative need [14], or communication efficiency [16], whereby the differentiated color terminology enables greater chromatic precision of the conveyed color-term denotata. The end-point of categorical and lexical differentiation of color space is, however, unclear.

New Developments in Understanding of Color Category “Basicness”

The challenges to the notion of the 11 BCC/BCT ceiling triggered fresh inquiries into the criteria of the color term “basicness.” Berlin and Kay [6] introduced three main and three auxiliary criteria of “basicness.” Paraphrasing these, to qualify as a BCT, a color name should be:

1. Understood and widely used by all speakers of a language
2. Generalizable across semantic domains and be applicable to any object
3. Psychologically salient
4. Reliably distinguishable from its neighbors in color space

The first two (1, 2) linguistic criteria more recently were scrutinized by Kerttula [19] in her analysis of the British English Corpus: by assessing primacy of color terms (criterion 2), their frequency, number of referents, and derivational productivity and by combining scores for each of these parameters, she arrived at the conclusion of “relative basicness” or a continuum of “basicness” between the most and the least basic color terms. Further, it was found that the ranking matches the order, in which English BCTs have developed historically, whereby a basic color-term status accumulates over time.

The last two criteria (3, 4) are behavioral and emerged more prominently in recent computational studies. Assessment of color term’s psychological saliency (criterion 3) newly underwent conceptual reconsideration: compared to the

broadly accepted conception of psychological saliency as the index derived from the term's order of occurrence and frequency in a color-name listing task (i.e., determined linguistically), alternative notions do not rely on color lexicons. In particular, using WCS color-naming data, Conway et al. [16] cognize psychological saliency of a BCT via the index of its communicative efficiency or surprisal; in comparison, Fider et al. [20] estimate BCT's psychological saliency as the degree of naming consensus by the population members, measured by its "color strength." Finally, criterion (4) is probed in psychophysical studies that employ color stimuli meticulously specified in terms of just noticeable differences (JNDs), to gauge performance at the boundaries of BCCs, in order to capture the influence that language labels exert over the perception and formation of boundaries (e.g., [21] (► "Dynamics of Color Category Formation and Boundaries")).

Visualization of Color Category Concepts

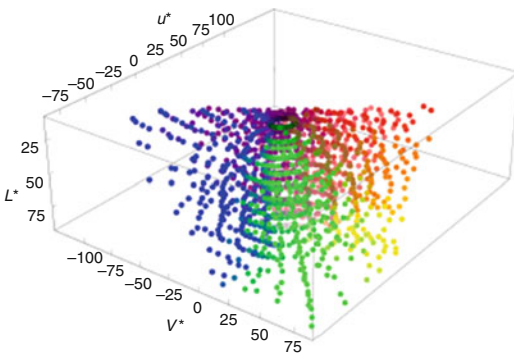
A recent development, driven by empirical studies, computational modeling, and epistemology, is a geometric approach to visualize BCCs, represented as regions of denotata that optimally partition color space (Fig. 1) [22]. The regions of the conceptual color space are "well-formed" (convex), i.e., the colors falling under any BCC

are maximally similar to each other and maximally dissimilar to colors falling under other BCCs (cf. [10]). Such mental representation of color categories results from efficient communication: it reflects an agreement among language speakers in naming colors and is informative and parsimonious, i.e., does not overload memory [22].

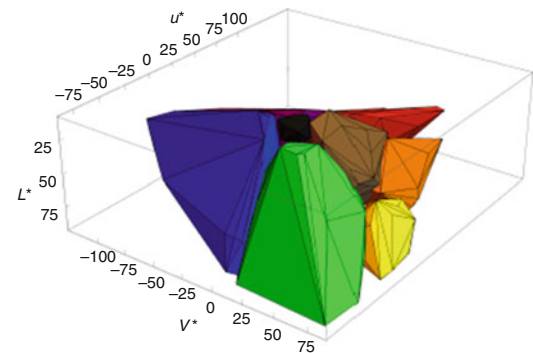
Color categories rendered as convex hulls (CIELAB space) enable to represent the structure of individual color spaces and, thus, to capture interindividual differences in color-naming patterns, as well as diatopic variations of color categorization within the same language (e.g., [23]). This line of research opens an avenue for quantifying differences in color categorization patterns between languages, to enable adequate rendition of color-term meanings between these. In addition, it can help apprehending dynamic changes of conceptual color space in ontogenesis and those resulting from bi-/multilingualism.

Concluding Remarks

The field of color categorization currently undergoes further intensive research enquiry – by investigating its universal mechanisms, pattern, structure, and dynamics, determined by perceptual constraints and communication pressure, as well as modulation by culture-specific factors. The ways of research enquiry greatly expanded



Color Categorization: Patterns and Mechanisms of Evolution, Fig. 1 The map of color stimuli densely representing color space (left) partitioned into 11 English basic color categories (right) presented as convex hulls in



CIELUV space. Fig. 4 reprinted from Cognition **193**:104007, Douven, I., Putting prototypes in place. doi: <https://doi.org/10.1016/j.cognition.2019.104007> [22], Elsevier Copyright Clearance Center (2019)

recently: along with continuing empirical studies – of color naming using methods of linguistics and psycholinguistics and gauging color similarity using psychophysical methods – there is a surge of studies that apply elaborate mathematical modeling and computational simulations, all of these providing new insights into the “color concept” mind. Beyond this, color categorization and naming research became the vehicle probing the workings of the mind in general, in particular, the Sapir-Whorf relativity hypothesis of the language effects on perception, and better understanding of the role that sociocultural transmission plays in generating shared linguistic categories among individuals for efficient exchange about the stimuli perceived in the environment

Cross-References

- [Bayesian Approaches to Color Category Learning](#)
- [Berlin and Kay Theory](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Categorical Perception](#)
- [Color Categories and Vagueness](#)
- [Color Category Learning in Naming-Game Simulations](#)
- [Color Dictionaries and Corpora](#)
- [Cross-cultural Communication](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Mesoamerican Color Survey Digital Archive](#)
- [Psychological Color Space and Color Terms](#)
- [World Color Survey](#)

References

1. Kay, P.: Chapter 11: Universality of color categorization. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) *Handbook of Color Psychology*, pp. 245–258. Cambridge University Press, Cambridge, UK (2015)
2. Davidoff, J.: Chapter 12: Color categorization across cultures. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) *Handbook of Color Psychology*, pp. 259–278. Cambridge University Press, Cambridge, UK (2015)
3. Witzel, C.: Misconceptions about colour categories. *Rev. Phil. Psych.* **10**, 499–540 (2019). <https://doi.org/10.1007/s13164-018-0404-5>
4. Dedrick, D.: *Naming the Rainbow: Colour Language, Colour Science, and Culture*. Kluwer, Dordrecht (1998)
5. Saunders, B.A.C., van Brakel, J.: Are there non-trivial constraints on colour categorisation? *Behav. Brain Sci.* **20**, 167–179 (1997)
6. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley/Los Angeles (1969/1991)
7. Douven, I., Wenmackers, S., Jraissati, Y., Decock, L.: Measuring graded membership: the case of color. *Cog. Sci.* **41**, 686–722 (2017)
8. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009). <http://www.icsi.berkeley.edu/wcs/data.html>
9. Lindsey, D.T., Brown, A.M.: Chapter 36: Color appearance, language, and neural coding. In: Werner, J.S., Chalupa, L.M. (eds.) *The New Visual Neurosciences*, pp. 511–531. MIT Press, Cambridge, MA (2014)
10. Jameson, K.A., D'Andrade, R.G.: Chapter 14: It's not really red, green, yellow, blue: an inquiry into cognitive color space. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 295–318. Cambridge University Press, Cambridge, MA (1997)
11. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modeling*, pp. 29–53. John Benjamins, Amsterdam/Philadelphia (2007)
12. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
13. Roberson, D., Davidoff, J.B., Davies, I.R.L., Shapiro, L.R.: Color categories: evidence for the cultural relativity hypothesis. *Cog. Psychol.* **50**, 378–411 (2005)
14. Regier, T., Carstensen, A., Kemp, C.: Languages support efficient communication about the environment: words for snow revisited. *PLoS ONE* **11**(4), e0151138. <https://doi.org/10.1371/journal.pone.0151138>
15. Paramei, G.V., Bimler, D.L.: Chapter 6: Language and psychology in the Modern Age. In: Steinvall, A., Street, S. (eds.) *Cultural History of Color, Vol. 6: The Modern Age*. Bloomsbury, London (2020). (in press)
16. Conway, B.R., Ratnasingam, S., Jara-E, Futrell, R., Gibson, E.: Communication efficiency of color naming

across languages provides a new framework for the evolution of color terms. *Cognition* **195**:104086 (2020). <https://doi.org/10.1016/j.cognition.2019.104086>

17. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language: a case study for colour. *Behav. Brain Sci.* **28**, 469–489 (2005)
18. Jameson, K.A., Komarova, N.L.: Evolutionary models of color categorization, II. Realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A.* **26**, 1424–1436 (2009)
19. Kerttula, S.: English Colour Terms: Etymology, Chronology, and Relative Basicness. *Mémoires de la Société Néophilologique de Helsinki*, Vol. LX, Société Néophilologique, Helsinki (2002)
20. Fider, N., Narens, L., Jameson, K.A., Komarova, N.L.: Quantitative approach for defining basic color terms and color category best exemplars. *J. Opt. Soc. Am. A.* **34**, 1285–1300 (2017)
21. Martinovic, J., Paramei, G.V., MacInnes, W.J.: Russian blues reveal the limits of language influencing colour discrimination. *Cognition* **201**:104281 (2020). <https://doi.org/10.1016/j.cognition.2020.104281>
22. Douven, I.: Putting prototypes in place. *Cognition* **193**:104007 (2019). <https://doi.org/10.1016/j.cognition.2019.104007>
23. Paramei, G.V., D’Orsi, M., Menegaz, G.: Chapter 5: Diatopic variation in the referential meaning of the “Italian blues”. In: MacDonald, L. W., Biggam, C.P., Paramei, G.V. (eds.) *Progress in Colour Studies: Cognition, Language and Beyond*, pp. 83–106. John Benjamins, Amsterdam/Philadelphia (2018)

Color Category Boundaries

► Color Name Applications in Computer Vision

Color Category Learning in Naming-Game Simulations

Tony Belpaeme
School of Computing and Mathematics,
University of Plymouth, Plymouth, UK

Synonyms

Computer modeling; Linguistic relativism; Simulation

Definition

Color categories are inextricably linked to language as color categories are typically, though not necessarily, associated with color terms. It is believed that the acquisition of categories, including color categories, is influenced by language [1]. Prelinguistic infants do seem to have a set of color categories, which are then either consolidated or modified through observing or engaging in linguistic interactions about color.

Language has been shown to have an influence on a range of modalities, such as time, space, and color. This phenomenon, known as linguistic relativism, shows how the language one uses, and by extension the culture one lives in, has an impact on perception and cognition. It has also been shown that language has an influence on color perception: having a particular color word speeds up spotting a chip of that color among distracting color chips [2, 3]. What is not entirely clear is how language influences the acquisition of color categories. As data on color category acquisition in infants is hard to come by, we can resort to computer simulations to learn more about how language impacts the acquisition of color categories.

A language is a communication system that is shared within a group of language users. As such, a language can be seen as an agreement between all language users on the words and rules of a language, and their meaning. Color words are also subject to this agreement: in the English language speakers agree to use “red” for, among others, the chromatic perception of a ripe tomato, blood, and a light with dominant spectral wavelength of 780 nm. There is no central authority insisting on this: language users themselves agree on this convention. When a new language user, such as a newborn child, enters a linguistic community, it will to varying degrees adopt this convention.

If language is a convention that is agreed upon by a linguistic community, and if language impacts category acquisition, then it follows that categories are to a certain extent also agreed upon by the community. Computer simulations can help us understand how a community can arrive

at an agreement on linguistic conventions and how language shapes concepts and categories.

There are a number of simulation models that can elucidate the process of language acquisition. The Language Game model [4, 5] has proven to be both effective and popular and can be used to study how a group of individuals reach a consensus on linguistic forms and associated categories. If, for example, a new word is introduced in a language, the Language Game model can simulate how that word spreads through the population. The model as such serves to study the dynamics of language change, and by setting parameters of the model, one can study what conditions make a language change. Iterated Learning Models, an alternative model to Language Games, study the sequential transfer of language [6]. Individuals are placed in a chain, and each individual's output is used as input for the next individual. Language Games study horizontal transmission, while Iterated Learning Models study vertical transmission of language. Both models demonstrate how small biases and communication bottlenecks can have a large effect on the language and conceptual structures that arise. A third class of simulated Language Game models are based on Evolutionary Game Theory (e.g., [7, 8]) or Statistical Mechanics (e.g., [9]). These are in essence mathematical models which start from a minimal set of parameters and study the influence of different settings of these parameters.

Language Game Models

In Language Game models, a community of language users is modeled as a population of N software agents. Each agent can store and recall words (or other linguistic information, such as rules) and categories. In the domain of color, the agents store color terms and color categories. In addition, each agent stores associations between color terms and color categories. An association typically is a value showing how strong the association between a term and category is. Agents start with empty inventories and gradually fill these with words and categories.

Various Language Game models represent color in different ways. Color can be modeled as a point on a single circular dimension [7, 8, 10, 11]; a color stimulus is then a real number in the interval $[0, 1]$. For a more realistic model of color, one can endow the artificial agents with a color appearance model, such as the CIE $L^*a^*b^*$ color space [12]. In this space, each color is represented by three real numbers L^* , a^* , and b^* , with L^* representing lightness, a^* the amount of green or magenta, and b^* the amount of yellow and blue. The CIE $L^*a^*b^*$ color appearance model aims to provide an accurate representation of color perception differences and allows for a similarity measure to be calculated between two colors, which is done by taking the Euclidean distance between two color values, permitting a good first approximation to categorical color perception (see [13] for an experimental appraisal and extension of CIE $L^*a^*b^*$).

In addition to the color categories and color terms used by an agent, simulations also need to prescribe what agents do when interacting with each other. In one form of a Language Game model interaction two agents are selected at random from the population; one acts as a *speaker*, the other as a *hearer*. Both agents are presented with a *context*; this is a set of M random color stimuli, each at a distance d from each other. The distance d guarantees that colors are not too similar or identical. From the context one color stimulus is selected, this will be the topic, and the speaker will attempt to communicate what the topic is to the hearer.

The speaker first finds a category that best matches the topic (often, but necessarily, this category is a unique match, meaning that it matches no other stimuli in the context). If no category can be found, the speaker will adapt its category set by adding a new category. Next it finds a word associated with the category and communicates it to the hearer. The hearer will attempt to guess the topic by looking up the word and the associated category in its inventory. It will check which stimulus matches the category best and will “point out” the stimulus. The hearer will then signal if this stimulus is indeed the intended topic. If it is, the game is successful. If the hearer

points out any other stimulus, the game fails. When successful, categories and word-category associations in both agents are reinforced, with the categories used in the interaction adapting such that they match the topic closer. When the game fails, the associations are weakened [4, 5].

During the iterative playing of Language Games [6] the agents create novel categories and words and change existing categories and category-word associations to optimize the communication. Only when communication is good enough (as determined by a preset threshold, e.g., $\tau = 90\%$ of all games end in success) will the agents' internal representations stop changing. It is important to note that the internal representation of all agents at this moment will not be identical and that the agents do not necessarily have the same number of words and categories. Their internal representations are merely sufficiently coordinated for communication to succeed with a success rate of τ . A population typically has a size N between two and several thousand agents and will play tens of thousands or perhaps millions of Language Games before stabilizing. An interesting observation is that the lexical and category systems of the agents adapt until they are "good enough" to successfully communicate; the agents do not need identical words and identical categories; they only need to sufficiently overlap to allow successful communication in most interactions. As such the semantics of words differ between agents: indeed your red is different from my red. As such, word and category are not true descriptions of the world; they are merely useful [14].

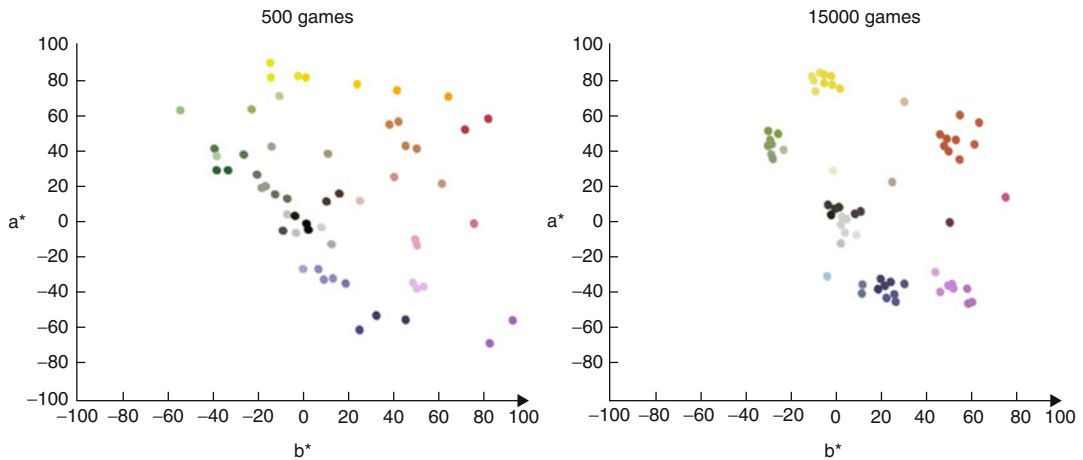
The dynamics of Language Game models have been extensively studied, as they inform research into the diachronic evolution of language. For example, the conditions under which a new word or a new linguistic construction is taken up by a language community can be modeled using Language Games [15, 16], and model predictions have been confirmed in studies with human participants [17]. Language Games have been used to clarify the minimum constraints needed to evolve a shared color category system by populations of agents [7], how varying agent perceptual properties impact color category system evolution [18–

20], and how varying color salience affects color category evolution [20]. In the case of color, however, the Language Game model serves a different purpose. It helps us to understand how relatively small biases present in color communications can have large-scale effects on the evolution of color category systems. Small biases are amplified through repeated interactions between language users. Specifically, simulated Language Game models help us formally investigate factors likely to influence color communications and help us understand why color categories appear to be universal and the degree to which pragmatics of communication or culture may contribute to color category evolution.

Explaining the Universal Character of Color Categories

It has been suggested that human color categories exhibit a universal pattern: many cultures have color categories that are seemingly similar. This was first suggested based on tenuous evidence in 1969 by Berlin and Kay [21] and later refined in the World Color Survey [22–24]. As such, color categories are not arbitrary, and this infused a principled search for the basis of their universal character. There are cultures which deviate sufficiently from the universal pattern, virtually ruling out the possibility that color categories are genetically determined. Other processes must be at work, and computer simulations can help us elucidate these.

When agents in a Language Game are communicating about color stimuli, their categories soon converge on a limited set of categories (Fig. 1). The locations in which the categories converge are not predetermined; instead they result from the slight biases introduced by the perceptual system of the agents [12, 18–20, 25]. Two different runs of a simulation will result in two different end results. However, repeated simulations do exhibit a pattern in which some color categories consistently emerge from the interaction between the agents. Categories such as red, green, yellow, blue, white, black, and so forth tend to almost always emerge. This matches the observations in



Color Category Learning in Naming-Game Simulations, Fig. 1 Both plots show all color categories of ten agents plotted in the a^*b^* plane of the CIE $L^*a^*b^*$ color space (the L^* lightness dimension is projected onto the a^*b^* plane; as such *white* and *black* categories are plotted on *top* of each other). The *left* figure shows the state after

500 Naming Games; the categories are still scattered in the color space. The *right* figure shows the state after 15,000 games when the communicate success of the agents is over 85%; the categories have gravitated to a limited number of clusters

the World Color Survey: these basic color terms and categories are found in the majority of the world's languages.

As the repeated playing of Language Games forces agents to coordinate their color categories and color terms, small biases in the agents' color perception will have a large influence. Belpaeme, Bleys and Steels [12, 23] showed how the bias of the CIE $L^*a^*b^*$ color appearance model together with a repeated negotiation of linguistic conventions results in the emergence of universal patterns of color categories. Baronchelli et al. [11] refined this; again using a Language Game model they showed how the human Just Noticeable Difference (JND) function, a function which shows the wavelength differences that are just about distinguishable to the human eye for each wavelength in the visual spectrum, also provides a small but important bias that can explain the universal character of color categories.

Language Games show how a variety of factors may contribute to the universal character of color categories without the need for color categories to be explicitly genetically determined. They permit the evaluation of, for example, neurophysical properties of human color perception as well as other small biases which, through

repeated linguistic negotiations, amplify and can contribute to the similarities seen across groups of languages that have roughly similar color categories.

References

1. Bowerman, M., Levinson, S. (eds.): Language Acquisition and Conceptual Development. Cambridge University Press, Cambridge (2001)
2. Gilbert, A., Regier, T., Kay, P., Ivry, R.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci.* **103**(2), 489–494 (2006)
3. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of colour in the left and right visual field is verbally mediated: evidence from Korean. *Cognition*. **107**(2), 752–762 (2008)
4. Steels, L.: Evolving grounded communication for robots. *Trends Cogn. Sci.* **7**(7), 308–312 (2003)
5. Steels, L.: Modeling the cultural evolution of language. *Phys. Life Rev.* **8**(4), 339–356 (2011)
6. Smith, K., Kirby, S., Brighton, H.: Iterated learning: a framework for the emergence of language. *Artif. Life*. **9**(4), 371–386 (2003)
7. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)
8. Steingrimsson, R.: Evolutionary game theoretical model of the evolution of the concept of hue, a hue structure, and color categorization in novice and stable

- learners. *Adv. Complex Syst.* **15**(03n04), 1150018 (2012)
9. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591 (2009)
 10. Puglisi, A., Baronchelli, A., Loreto, V.: Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci.* **105**(23), 7936–7940 (2008)
 11. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci.* **107**, 2403 (2010)
 12. Belpaeme, T., Bleys, J.: Explaining universal colour categories through a constrained acquisition process. *Adapt. Behav.* **13**(4), 293–310 (2005)
 13. Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. *PLoS One.* **8**, e55986 (2013)
 14. Hoffman, D., Singh, M.: Computational evolutionary perception. *Perception.* **41**, 1073–1091 (2012) (Special issue in honor of David Marr)
 15. de Vylder, B., Tuyls, K.: How to reach linguistic consensus: a proof of convergence for the naming game. *J. Theor. Biol.* **242**(4), 818–831 (2006)
 16. Loreto, V., Steels, L.: Social dynamics: emergence of language. *Nat. Phys.* **3**(11), 758–760 (2007)
 17. Centola, D., Baronchelli, A.: The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl. Acad. Sci.* **112**(7), 1989–1994 (2015)
 18. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am. A.* **26**(6), 1414–1423 (2009)
 19. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A.* **26**(6), 1424–1436 (2009)
 20. Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *J. Theor. Biol.* **253**(4), 680–700 (2008)
 21. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution.* University of California Press, Berkeley (1969)
 22. Kay, P., Berlin, B., Maffi, L., Merrifield, W.: *The World Color Survey.* Center for the Study of Language and Information, Stanford (2003)
 23. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci.* **100**(15), 9085–9089 (2003)
 24. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within – language diversity. *Proc. Natl. Acad. Sci.* **106**(47), 19785–19790 (2009)
 25. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language. A case study for colour. *Behav. Brain Sci.* **24**(8), 469–529 (2005)

Color Centers

Richard J. D. Tilley
Queen's Buildings, Cardiff University,
Cardiff, UK

Definition

Color centers are point defects or point defect clusters associated with trapped electrons or holes in normally transparent materials. These centers cause the solid to become colored when the electronic ground state of the defect is excited to higher energy states by the absorption of visible light [1–5]. [Note that transition metal and lanthanoid ion dopants that engender color in an otherwise colorless matrix are frequently called color centers. These are dealt with elsewhere (see “[Cross-References](#)”).]

Color Centers

The concept of color arising from point defects was initially developed in the first half of the twentieth century, principally by Pohl, in Germany. It was discovered that clear alkali halide crystals could be made intensely colored by diverse methods, including irradiation by X-rays, heating crystals in the vapor of any alkali metal, and electrolysis. Crystals with induced color were found to have a lower density than the crystals before treatment and appeared to contain a population of anion vacancies. The absorption spectrum was always a simple bell shape. It was notable that the color engendered in the crystal was always the same and was not dependent upon the method of color production. That is, if a crystal of KCl was heated in an atmosphere of any alkali metal, or irradiated by X-rays, or electrolyzed using any cathode material, the crystal took on a violet color. Similarly, crystals of NaCl always took on an orange brown hue under all preparation methods. The fact that the color was a unique property of the host structure implied that it was a property of the crystal itself. The color was

ultimately attributed to the formation of defects called *Farbzentren* (= color centers), which were equated with mistakes in the crystal structure.

Since these early studies, many different color center types have been characterized (Table 1). All are electronic defects that possess similar characteristics in that they form in colorless, insulating, often relatively ionic, solids and incorporate either trapped electrons, to produce *electron excess centers*, or trapped holes, to produce *hole excess centers*. Solids may contain several different types of color center, including populations of both electron and hole excess centers in the same host matrix. Color centers are usually labelled with a letter symbol (Table 1).

As the alkali halide studies demonstrated, color centers can be created in a host matrix in a number of ways. For example, strong ultraviolet light can transform clear glass into purple-colored “desert glass” and intense radiation from nuclear weapons or accidents may color ceramics such as porcelain a blue color, both changes being due to the formation of color centers. Controlled irradiation in nuclear reactors or similar is used to produce artificial gemstones from colorless and less

valuable starting materials. For example, irradiation of colorless topaz, $\text{Al}_2\text{SiO}_4(\text{F}, \text{OH})_2$, induces a beautiful blue color due to color center formation (Fig. 1).

From an optical viewpoint, color centers behave something like isolated atoms dispersed throughout the host matrix. As such, they make ideal probes of the interaction of light with the matrix and are currently being explored for wide ranging electro-optic applications.

F Centers: Electron Excess Centers

The first color center to be characterized was the F center found in alkali halides, and these remain the best-known electron excess centers. Anion vacancies, which occur in low concentrations in alkali halides, MX , have an effective positive charge and an F center is an anion vacancy (V_X) plus a trapped electron to form an effectively neutral defect ($\text{V}_\text{X} \text{e}$) (Fig. 2a). The F center behaves rather like an isolated hydrogen atom in the structure, and the electron is able to absorb electromagnetic radiation, jumping from one energy level to another, just as an electron absorbs radiation in the Bohr model of the H atom. The

Color Centers, Table 1 Some color centers

Host crystal	Symbol	Description
Alkali halide: MX	F	Electron trapped an anion vacancy
	M, F_2	A pair of adjacent interacting F centers
	F_A	F center next to an alkaline metal substitutional impurity
	F', F^-	F center with two trapped electrons
	$\text{R}, \text{M}^+, \text{F}_2^+$	Three adjacent interacting F centers on (111)
	V_K	Two adjacent anion vacancies with one trapped electron
Alkaline earth halide: MX_2	F	Electron trapped an anion vacancy
	M	A pair of adjacent F centers aligned along [100]
	F_3	Three adjacent F centers aligned along [100]
Alkaline earth oxide: MO	F	Oxygen vacancy with two trapped electrons
	F', F^-	Oxygen vacancy with three trapped electrons
	F^+	Oxygen vacancy with one trapped electron
Quartz: SiO_2	E', E^-	Oxygen vacancy with one trapped electron
Diamond: C	C, P1	Isolated N atom substituted for a C atom
	A	Two C atoms substituted by N, forming an N-N pair
	N3	Three N atoms on C sites surrounding a C vacancy
	N2	Two N atoms on C sites adjacent to a C vacancy
	NV	One N atom on a carbon site adjacent to a C vacancy
	NV^-	Negatively charged NV center

peak of the absorption curve, $\lambda_{\max}$, corresponds to a total removal of the electron from the F center, and usually falls in the visible, so coloring the originally transparent crystals (Table 2). In terms of band theory, alkali halides are insulators with a considerable energy gap between a filled valence band and an empty conduction band. The F center in its ground state creates a new energy level or narrow energy band in the band gap (Fig. 2b). The



Color Centers, Fig. 1 Blue topaz, produced by irradiation of clear crystals; the color arises from a population of color centers. (Photograph R J D Tilley)

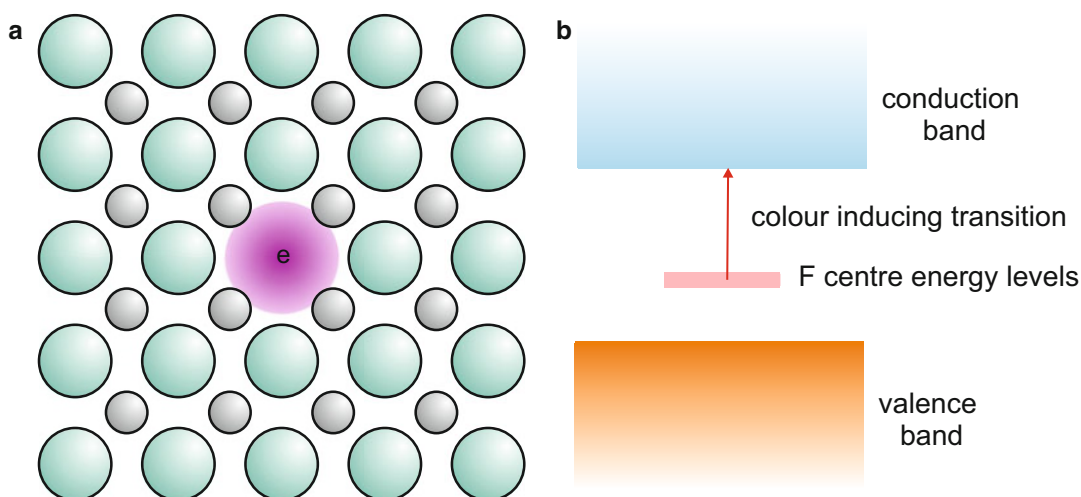
color-producing optical absorption peak corresponds to electron promotion into the conduction band. The variation in the color observed depends upon the host crystal band gap and the energy level of the F center, both of which are linked to the lattice parameter of the host matrix (Table 2).

The detailed mechanism for the formation of F centers depends to some extent upon the manner in which they are generated. In the case of irradiation by X-rays, for example, the energetic radiation is able to displace an electron from a normal anion, and some of these become trapped at existing anion vacancies. The corresponding anion that has lost an electron creates a hole energy level in the valence band.

F centers occur in many alkaline earth halides and oxides (Table 1). For example, the mineral *Blue John* is a rare naturally occurring purple blue form of fluorite, CaF_2 . The coloration is caused by F centers believed to have formed when the fluorite crystals were fortuitously located near to uranium compounds in the rock strata. Radioactive decay of the uranium produced the energetic radiation necessary to form the color centers.

Hole Excess Centers

One of the best understood hole excess centers gives rise to the color in smoky quartz, a naturally occurring form of silica (SiO_2). This material



Color Centers, Fig. 2 (a) An F center in an alkali halide MX crystal (schematic); large circles, anions X^- , small circles, cations, M^+ ; (b) schematic energy level diagram for an F center in an alkali halide crystal

Color Centers, Table 2 Alkali metal halide F centers

Compound	Absorption peak $\lambda_{\text{max}}/\text{nm}$	Color ^a	Lattice parameter/nm
LiF	235, ultraviolet	Colorless	0.4073
NaF	345, ultraviolet	Colorless	0.4620
KF	460, blue	Yellow brown	0.5347
RbF	510, green	Magenta	0.5640
LiCl	390, ultraviolet (just)	Yellow green	0.5130
NaCl	460, blue	Yellow brown	0.5641
KCl	565, green	Violet	0.6293
RbCl	620, orange	Blue green	0.6581
LiBr	460, blue	Yellow brown	0.5501
NaBr	540, green	Purple	0.5973
KBr	620, orange	Blue green	0.6600
RbBr	690, red	Blue green	0.6854

^aThe appearance of the crystal is the complementary color to that removed by the absorption band

contains small amounts of Al^{3+} substitutional impurities. These replace Si^{4+} ions in $[\text{SiO}_4]$ tetrahedral units which form the building units of the crystal. Overall charge neutrality is preserved by the incorporation of one H^+ for each Al^{3+} . These H^+ ions sit in interstitial positions in the rather open SiO_2 structure. The color center giving rise to the smoky color in quartz is formed when an electron is liberated from an $[\text{AlO}_4]$ unit by ionizing radiation and is trapped on one of the H^+ ions present, leaving a hole (h) behind. The color center is a $(\text{AlO}_4 \text{ h})$ group. The color arises when the trapped hole absorbs radiation exactly as the electron in an F center.

The situation in amethyst, which is a form of silica containing Fe^{3+} and H^+ impurities, is similar. The impurity Fe^{3+} forms $[\text{FeO}_4]$ groups. These crystals are known as the pale yellow semiprecious gemstone citrine and also in a pale green form; the color arising from the impurity Fe^{3+} ions (see “[Cross-References](#)”). On irradiation, $(\text{FeO}_4 \text{ h})$ color centers form by interaction with H^+ ions. The color centers impart the purple amethyst coloration to the crystals (Fig. 3).

Color Centers in Diamond

Colored Diamonds

The diamond structure is built up of carbon atoms each coordinated to four carbon atom neighbors;



Color Centers, Fig. 3 Amethyst crystals; the purple color arises from a population of color centers. The pale green stone is unirradiated. (Photograph R J D Tilley)

the linking being via tetrahedral sp^3 hybrid bonds (Fig. 4a). Diamond has a band gap of about 5.5 eV which is too large to absorb visible light and perfect diamonds are clear. The commonest impurity in natural diamonds is nitrogen, which mostly substitutes for carbon on normal tetrahedral sites in the crystal. Natural diamonds are often subjected to temperatures of 1000–1200 °C, over geological timescales, allowing these nitrogen atoms to diffuse through the structure, leading to

populations of defect clusters as well as isolated point defects.

The color of the highly prized natural yellow diamonds called *Canaries* is due to isolated nitrogen atoms located on carbon sites, which form color centers called C centers. The color arises in the following way. Nitrogen, with an electron configuration $1s^2 2s^2 2p^3$, has five bonding electrons, one more than carbon, with a configuration $1s^2 2s^2 2p^2$. Four of the electrons around each impurity nitrogen atom are used to fulfill the local sp^3 bonding requirements of the crystal structure and one electron remains unused. Substitution of nitrogen for carbon on a normal carbon atom site in the crystal thus creates an electron excess color center (Fig. 4b). On an energy level diagram, this gives rise to a donor level in the band gap, which, because of lattice vibrations and other interactions, consists of a narrow band of energies centered at 2.2 eV and extending to 1.7 eV, the ionization energy of the N atom in diamond. The electron can be excited into the conduction band by absorption of incident visible light of wavelengths longer than about 564 nm, giving the

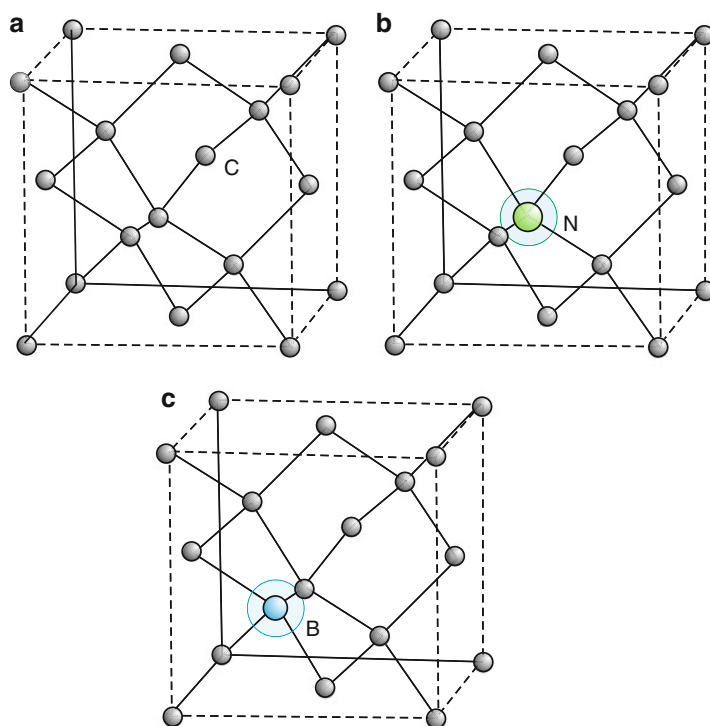
stones a faint yellow aspect. As the number of C centers increases, the color intensifies.

The N3 center, which consists of three nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy, seems to be responsible (at least in part) for the pale straw color of expensive *Cape Yellow* diamonds. The N3 center has a complex electronic structure which absorbs just in the blue end of the visible, at 415 nm, resulting in yellowish stones. The N3 centers are often accompanied by neutral N2 centers consisting of two nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy. These absorb at approximately 475 nm, giving a yellow color to the stones and adding to that contributed by the N3 clusters. When crystals are irradiated, either naturally or artificially, the N2 cluster can trap an electron to form a negatively charged $N2^-$ center that has an absorption peak at approximately 989 nm in the infrared. This absorption band can spill over into the red part of the visible spectrum, leading to stones with a blue tone and producing blue diamonds. When all these nitrogen-based color

C

Color Centers, Fig. 4

(a) The diamond structure as a linkage of tetrahedrally coordinated carbon atoms (*small circles*); (b) structure of a C center in diamond consisting of a nitrogen atom (*large circle*) occupying a carbon position, schematic; (c) structure of a boron impurity center (*large circle*) in diamond (schematic)



centers are present in roughly equal quantities, the stones take on a green hue.

Although nitrogen-linked color centers give rise to the highly valued yellow hued diamonds, many prized blue diamonds are the result of boron impurities on normally occupied carbon atom sites (Fig. 4c). Boron, with an electron configuration $1s^2 2s^2 2p^1$, has only three outer bonding electrons instead of the four found on carbon. These three are used in fulfilling the bonding requirements of the structure, but one bond of the four is incomplete and lacks an electron, making the defect a hole excess center. In semiconductor physics terms, the center introduces a narrow band of energy levels approximately 0.4 eV above the valence band. The transition of an electron from the valence band to this band gives rise to an absorption peak in the infrared with a high-energy tail encroaching into the red at 700 nm. The boron-doped diamonds therefore absorb some red light and leave the gemstone with an overall blue color.

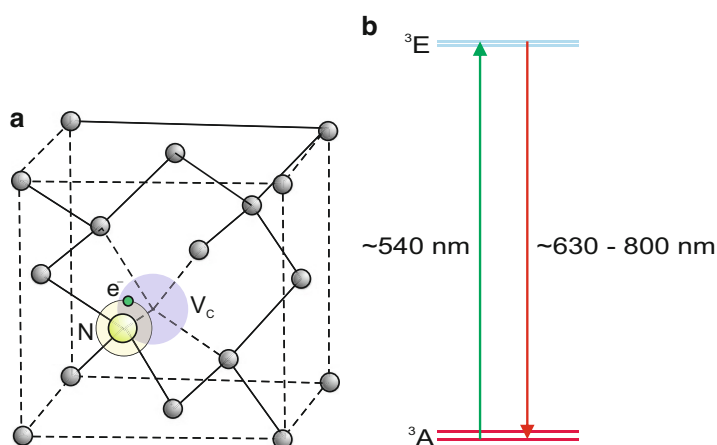
Nitrogen-Vacancy Centers

The diamond color center that has been studied in most detail is that consisting of a single nitrogen substitutional impurity located next to a carbon vacancy, together with a trapped electron to form a *negatively charged nitrogen–vacancy center*, NV^- (often just called an NV center). These defect centers are readily created by the irradiation of nitrogen containing natural or synthetic diamonds, diamond thin films or diamond

nanoparticles with high energy protons. The proton irradiation results in the formation of carbon vacancies, and if the crystals are then annealed at above 600 °C, the temperature at which the vacancies become mobile, they diffuse through the structure until they encounter a nitrogen impurity. The strain around the nitrogen atom effectively traps the vacancy, preventing further migration. In the resultant NV^- centers, the tetrahedron surrounding the carbon vacancy is composed of three carbon atoms and one nitrogen atom (Fig. 5a).

These centers are being investigated for applications, including room temperature quantum computing, nanoscale magnetometers, fluorescent markers, and biological imaging [6, 7]. The applications follow from the unique features of the energy levels of the (NV^-) center. The ground state term of the electronic structure is 3A and the first excited state term is 3E (Fig. 5b) [see “Cross-References”: Transition-Metal Ion Colors, for a description of 3A and 3E terminology.] (Note that both the ground state and excited state terms are split into several levels. This splitting is of prime importance for many applications but does not dominate the overall color aspects of the centers and can be ignored in the present context.) Excitation from the 3A ground state to the 3E excited state is by absorption of light over the range of approximately 514–560 nm, giving stones a pink hue. Emission falls in the range 630–800 nm, but the observed color is dominated by a particularly strong red fluorescence at

Color Centers, Fig. 5 (a) An NV^- center in diamond, schematic; (b) approximate energy level diagram of an NV^- center



637 nm. Under suitable observing conditions, single bright red fluorescent (NV^-) centers can be observed, making nanodiamonds that incorporate these defects ideal probes to track vital processes in living cells [7].

Tenebrescence

Tenebrescence (reversible photochromism) is the property of darkening, especially reversibly, in response to incident radiation. It is associated particularly with minerals which change color when exposed to sunlight. The color fades when the sunlight is shaded, and the effect can be repeated indefinitely. The best known minerals which exhibit tenebrescence are scapolite $(\text{Na,Ca})_4(\text{Al,Si})_3\text{Si}_6\text{O}_{24}(\text{Cl,CO}_3, \text{SO}_4)$; tugtupite, $\text{Na}_4\text{AlBeSi}_4\text{O}_{12}\text{Cl}$; and hackmanite, $\text{Na}_8\text{Al}_6\text{Si}_6\text{O}_{24}(\text{Cl,S})_2$, also given the formula $(\text{Na}_8(\text{Al,SiO}_4)_6(\text{Cl,S})_2$, reflecting that variable Al to Si concentrations may be present

Hackmanite is a derivative form of sodalite, $\text{Na}_8\text{Al}_6\text{Si}_6\text{O}_{24}\text{Cl}_2$, in which some of the Cl has been replaced by a sulfur containing group S_2^{2-} , SO_3^{2-} , or SO_4^{2-} . The tenebrescent colors observed in originally colorless hackmanite crystals range over pink, red, purple, and blue. These colors appear rapidly when the crystals are exposed to strong sunlight or ultraviolet light and then fade over the course of several minutes when in shadow. This coloring and fading process can be repeated indefinitely. For tenebrescence to occur it is found that only a low level of S is required, amounting to just a few percent. At greater concentrations, the effect is lost.

The structure of hackmanite is cubic and is built of rings of four apex-linked SiO_4 and AlO_4 tetrahedra centered on each cube face and rings of six similar apex-linked tetrahedra around each cell corner. These form cubo-octahedral cavities containing the Na^+ cations and the Cl^- anions. In order to maintain charge balance in the tenebrescent crystals, the interpolation of, for example, an S_2^{2-} anion into a cavity requires the loss of 2Cl^- anions, which thereby introduces an anion vacancy into a nearby cavity where it is associated with surrounding Na^+ cations in a

$(\text{Na}_4\text{V})^{4+}$ cluster. Although the mechanism for the tenebrescence has not been fully elucidated, it is believed that an electron is released from S_2^{2-} groups by ultraviolet irradiation and this migrates to the nearby $(\text{Na}_4\text{V})^{4+}$ cluster where it becomes trapped. This trapping center acts as a color center and so colors the normally clear crystals. The color observed is proportional to the lattice parameter of the hackmanite, which itself is a reflection of the size of the F center. The trap energy is quite low, however, and the electrons are liberated by thermal energy and diffuse back to the sulfur groups, where they are again incorporated. This leads to the diminishing color after the exciting radiation is removed.

Some Applications of Color Centers

F Center Lasers

F centers do not exhibit laser action but F centers that have a dopant cation next to the anion vacancy are used in this way. These are typified by F_{Li} centers, which consist of an F center with a lithium ion neighbor. Crystals of KCl or RbCl doped with LiCl, containing F_{Li} centers, have been found to be good laser materials, yielding emission lines with wavelengths between 2.45 and 3.45 μm . A unique property of these crystals is that in the excited state an anion adjacent to the F_{Li} center moves into an interstitial position. This is *type II* laser behavior, and the active centers are called F_{Li} (II) centers. These centers are stable if the crystal is kept at -10°C .

Persistent Luminescence

Color centers are active in materials that show persistent luminescence. In these compounds, irradiation by ultraviolet light present in normal daylight gives rise to luminescence for many hours after dark. Such materials are making their way into applications as diverse as road signs that do not need a power supply at night and bicycle frames that glow in the dark.

Although the color centers responsible for persistent luminescence vary from material to material, the principle can be illustrated with the oxide SrAl_2O_4 doped with B^{3+} , Eu^{2+} , and Dy^{3+} , which

gives a green luminescence. The SrAl_2O_4 structure consists of a framework of corner-linked $[\text{AlO}_4]$ tetrahedra enclosing Sr^{2+} ions in the cavities so formed. The B^{3+} substitutes for Al^{3+} to create $[\text{BO}_4]$ tetrahedra and $[\text{BO}_3]$ triangular groups, which can be thought of as $[\text{BO}_4]$ tetrahedra in which one of the oxygen ions is absent to form an oxygen vacancy (V_O), creating a $(\text{BO}_3 \text{V}_\text{O})$ center. The substitution of Dy^{3+} for Sr^{2+} results in a charge imbalance that is compensated for by the creation of an equal number of vacancies on Sr^{2+} positions (V_Sr) to form defect clusters ($\text{Dy BO}_4 \text{V}_\text{Sr}$). Following irradiation with ultraviolet light, an electron is transferred from a $(\text{BO}_3 \text{V}_\text{O})$ cluster to a $(\text{Dy BO}_4 \text{V}_\text{Sr})$ unit, resulting in the formation of two complex color centers: $(\text{Dy BO}_4 \text{V}_\text{Sr} \text{h})$ which are hole excess centers and $(\text{BO}_3 \text{V}_\text{O} \text{e})$, which are electron excess centers. The origin of the luminescence lies in this pair of color centers. Under normal conditions, the electron and hole centers are metastable, and over the course of several hours, the holes and electrons gradually recombine. The energy liberated is transferred to the Eu^{2+} ions, which lose energy by the emission of photons, thus producing a long-lasting green fluorescence.

Photostimulable Phosphors

Photostimulable phosphors that make use of color centers are widely used in X-ray imaging, particularly by dentists, where they have largely replaced X-ray film recording. The first commercial material to fulfill these requirements, introduced in 1983, was BaFBr doped with Eu^{2+} . Although the detailed mechanism by which these phosphors work is still not entirely clear, it is established that an important component of the process is the formation of F centers, produced as a result of the X-ray irradiation. In dental X-ray imaging, a plate covered with a thin layer of phosphor is placed into the mouth and exposed to X-rays. The X-rays initially displace an electron from an anion to form an electron–hole pair. The electron is subsequently trapped at an anion vacancy to form an F center. This fairly stable pair of electronic defects constitutes a *latent image* in the phosphor. Subsequently, the exposed plate is irradiated with 633 nm light from a helium–neon

laser. The electrons trapped in the F centers are initially promoted to the conduction band, after which they are free to recombine with the holes. The energy liberated is transferred to Eu^{2+} ions which decay from the subsequent excited state by the emission of visible light at 420 nm, subsequently recorded as a digital image. The number of F centers and holes, and therefore the amount of light emission, is proportional to the X-ray intensity in the phosphor. The optical image thus records accurately the degree to which the X-rays have penetrated the subject.

Postscript

The electronic structures of the neutral NV and negative NV^- color centers in diamond have been intensively studied, and it has taken some 35 years to reach the current understanding of energy levels of this defect. It would seem reasonable to suspect that similar detailed investigations of the other color centers described will also lead to significant revisions in their electronic structures and consequently a more precise description of their color-engendering abilities.

Cross-References

- [Lanthanoid Ion Color](#)
- [Luminescence](#)
- [Luminescent Materials](#)
- [Phosphors and Fluorescent Powders](#)
- [Transition Metal Ion Colors](#)

References

1. Tilley, R.J.D.: *Defects in Solids*. Wiley, Hoboken (2008)
2. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Hayes, W., Stoneham, A.M.: *Defects and Defect Processes in Nonmetallic Solids*. Wiley, New York (1985). Reprinted by Dover, Mineola (2004)
5. Fowler, W.B.: Chapter 2: The imperfect solid – color centers in ionic crystals. In: Hannay, N.B. (ed.) *Treatise*

- on Solid State Chemistry, Vol. 2, Defects in Solids. Plenum, New York (1975)
6. Acosta, V.: In: Hemmer, P. (ed.) Nitrogen-Vacancy Centres: Physics and Applications. M. R. S. Bulletin, vol. 38. Materials Research Society, Warrendale (2013)
 7. Wu, Y., Jelezko, F., Plenio, M.B., Weil, T.: Diamond quantum devices in biology. *Angew. Chem. Int. Ed.* **55**, 6586–6598 (2016)

Color Circle

Paul Green-Armytage
The School of Media, Creative Arts and Social
Enquiry, Curtin University, Perth, WA, Australia

Synonyms

[Color wheel](#); [Hue sequence](#)

Definition

The color circle, as generally understood and widely used, is a diagram with a continuous sequence of hues arranged in the order of the spectrum. (The gap between spectral red and spectral violet is bridged with extra-spectral purples.) The color circle diagram is used as a guide to mixing and color composition. It is also used in the classification of colors and is incorporated in all three-dimensional color order systems.

Introduction

Very many color circle diagrams have been designed and published. The essential feature is that the diagram must represent the sequence of hues in correct order and in a continuum: reds, oranges, yellows, greens, blues, purples, reds, oranges, yellows, greens, etc. The number of separate hues in the sequence can vary as can the hues themselves that are selected for inclusion. The starting point for some color circles is a choice of so-called primary colors. Other circles are organized so that colors opposite to each other are, in

some sense, complementary. Many color circles are also organized so that the degree of difference between neighboring colors in the sequence appears to be the same all round the circle.

The Variety of Color Circles

There is no single “correct” design for a color circle. Different circles have been constructed on different principles. They are not necessarily presented in color; in some the colors are simply identified by name. Color circles can be grouped in three broad categories: those that represent colors as something physical, those that represent colors as visual phenomena – what people see –, and those where the colors can be understood as either or both, physical and/or visual. In the first category the colors represent lights, paints, inks, or dyes, and the position of the colors in relationship to each other is generally determined by their physical properties or by the way that these lights, paints, inks, or dyes can be mixed to produce a large range of other colors. In the second category the colors are simply themselves and it is their appearance that determines how they are related in the circle. In the third category it may not be clear whether the colors are to be understood as physical or visual or both. It could be that the designers and users of such circles confuse or conflate the physical and the visual aspects of color. For an account of the way that the physical and visual aspects of color can be confused or conflated, see the entry on ► [“Primary Colors”](#) in this encyclopedia.

As color circles fall into different categories, and are constructed according to different principles, the relationship between colors can vary from circle to circle. This does not mean that the sequence of hues can vary – that must always be the same – but the distance between hues can vary. For example, it could be that two different circles have yellow in the 12:00 o’clock position, but red might be at 3:00 o’clock in one circle and at 4:00 o’clock in the other. The principles underlying the design of the color circles would have fixed the position of red in relation to yellow. Problems can arise when there is a mismatch between the intentions of the designer and the expectations of the

user. A particular color circle can be criticized for not representing particular color relationships when it was never the intention that such relationships be embodied in the design.

Precursors to the Color Circle

The color circle, as defined above, was invented by Isaac Newton, whose color circle is illustrated in his book *Opticks*, first published in 1704 (see Fig. 4) [1, pp. 32–35]. Before Newton it was believed that there was a beginning and an end to the sequence of colors. Aristotle had claimed that all colors derived from mixtures of white and black which he placed at either end of a linear scale [2, pp. 31–32]. Aristotle's prestige was so great that science was dominated by his ideas until the time of Galileo [3, p. 213]. Some circular color diagrams do predate Newton but they still reflect Aristotle's ideas.

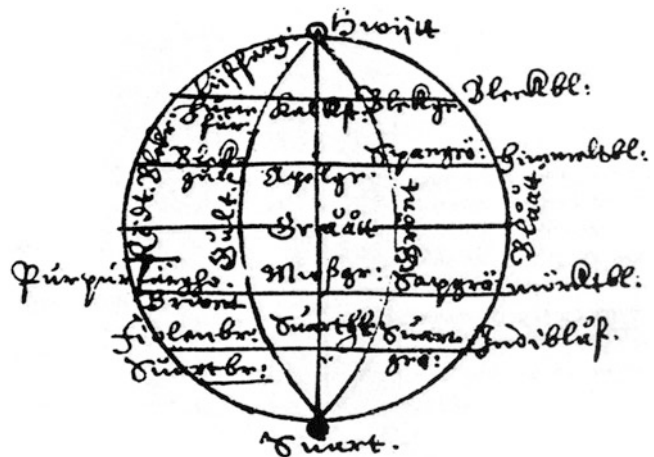
John Gage describes a circular diagram that survives in an illuminated manuscript from the fifteenth century. This shows the colors of urine, between white and black, that helped physicians diagnose certain diseases [4, pp. 162 and 171]. White and black also appear in other circular diagrams, two in a manuscript by Sigfrid Forsius (1611) and another in a printed medical text by Robert Fludd (1629–1631). Fludd's circle has this sequence: white, yellow, orange, red, green, blue, and black [2, p. 42]. Forsius echoes Aristotle

when he writes: “Among colours there are two principles, white and black, from which all others arise” [5, pp. 12–13]. The Forsius manuscript contains two circular diagrams. Each has white in the 12:00 o'clock position and black at 6:00 o'clock. Forsius explains that his first circle illustrates the way in which the “Ancients” arranged the colors. Down the left side, between white and black, are pale yellow, yellow, orange, red, purple, brown, and violet. Down the right side are ash gray, gray, sky blue, blue, pale green, green, and dark green. The second Forsius diagram is shown in Fig. 1 and recreated in Fig. 2 with color names translated from old Swedish.

Claims have been made that this second Forsius diagram represents a sphere [7, p. 224]. If this is correct then the central horizontal line must be read as a circle seen from the side with red, yellow, green, and blue on the circumference and gray in the middle. Werner Spillmann has applied the principles of graphic projection to this interpretation and points out that it would mean that the hues are in the wrong order [8, p. 7]. Yellow must be closest to the observer and green farthest away or vice versa. Either way, the hue sequence would be red, yellow, blue, and green, which is not the order in which they appear in the spectrum. However these early diagrams are read, not one of them, unequivocally, shows the sequence of hues as continuous and in the correct order.

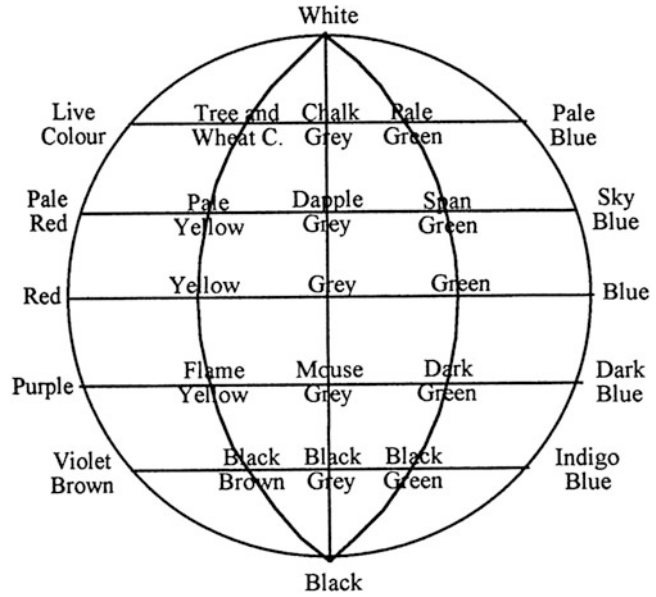
Color Circle,

Fig. 1 Circular diagram from the manuscript by Sigfrid Forsius. (Reproduced from Ref. [6], with permission)



Color Circle,

Fig. 2 Reconstruction of the Forsius diagram with color names translated from old Swedish



Development of the Color Circle

The color circle did not evolve so much as to develop a growing variety of uses. Newton’s circle identifies colors with light of different wavelengths and he shows how the diagram can be used to illustrate the results of additive mixing. Some later circles illustrate the results of subtractive mixing from a set of “primaries.” In other circles the positions of the colors are determined by how they appear. Such arrangements are used in systems of color identification. Color circles have been developed where the apparent difference between neighboring colors is the same all round the circle. Colors that are opposite to each other in some circles are described as being complementary. Even spacing and complementary relationships are seen as significant in theories of color harmony.

Newton

In Fig. 3 Lindsay MacDonald is showing how white light is refracted by a prism to reveal the different colors. Although not easily visible here, the blue shades into a slightly reddish violet at the short-wave end of the spectrum. Perhaps it was

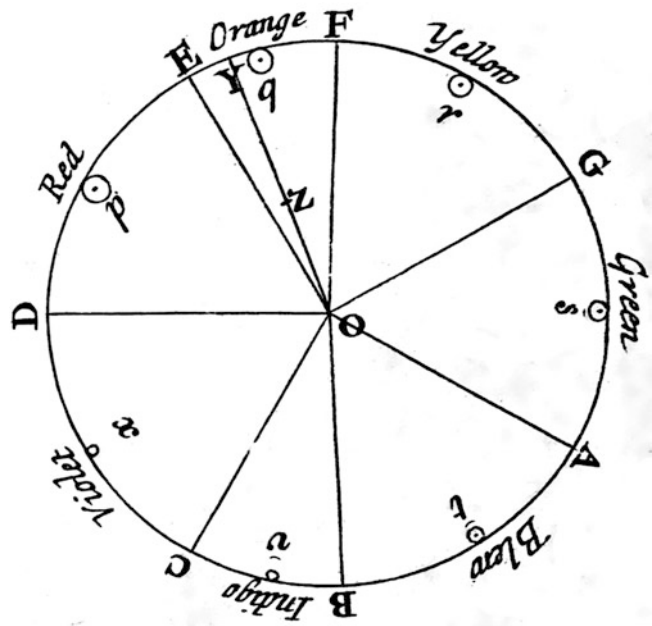


Color Circle, Fig. 3 Lindsay MacDonald demonstrating how a prism separates white light into different colors

the redness at each end of the spectrum that gave Newton the idea of connecting the two ends to form his circle. Newton’s color circle is shown in Fig. 4.

Strictly speaking, the sequence of hues in Newton’s circle is incomplete. The circle is divided into seven segments, each identified by name. At first Newton refers to each segment as representing a single color – “Let the first part

Color Circle, Fig. 4 Isaac Newton's color circle representing colors as different wavelengths of light. (Reproduced from Ref. [6], with permission)



DE represent a red, the second EF orange . . .” – but he goes on to say that these represent “all the colors of uncompounded light gradually passing into one another as they do when made by Prisms . . .” [1, p. 32]. So the segment DE does not represent a single red but a range of the hues that would be identified as “red.” And the lines that separate the segments represent the borders where colors that would be identified by one name give way to those that would be identified by the next; the “red” segment would have a range of colors from reds to orange-reds. Because the colors in the circle represent the wavelengths of light in the visible spectrum, there is no place for the bluish reds and purples that are not visible in the spectrum. And if the hues shade into each other across the borders separating the other named segments, there would be a break in the sequence at point D with no spectral hues to shade from violets to reds.

A color circle which does show all the hues, spectral and non-spectral, shading into each other was produced by Michel-Eugène Chevreul and published in 1864 (Fig. 5). This can be set beside another of Chevreul's circles (Fig. 6) which is divided into 72 hues. Chevreul's color circles are described by Verena Schindler [9, pp. 66–68].

The colors in Chevreul's circle between the 5:00 o'clock and 7:00 o'clock positions are not visible in the spectrum and so have no place in Newton's circle. However, if Newton had intended to represent seven hues only, with no shading from one named hue to the next, the gap between violet and red would be no more noticeable than that between red and orange. The hues are in the correct order and the sequence is continuous.

The First Color Circles Published in Color

The first color circles to be published in color appear in an enlarged edition of a book on miniature painting. The author of the first edition of 1673 has been identified as Claude Boutet, but the color circles were only added in the enlarged edition of 1708 in a new section on pastel painting. The unknown author of this later section writes: “Here are two circles by which one will be able to see how the primitive colors, yellow, fire red, crimson red and blue generate the other colors” [2, p. 57]. So these circles are demonstrations of subtractive mixing with paints (Fig. 7). The possibility of mixing a complete sequence of

Color Circle,

Fig. 5 Michel-Eugène Chevreul 1864. Color circle with hues shading into each other. (Reproduced from Ref. [6], with permission)



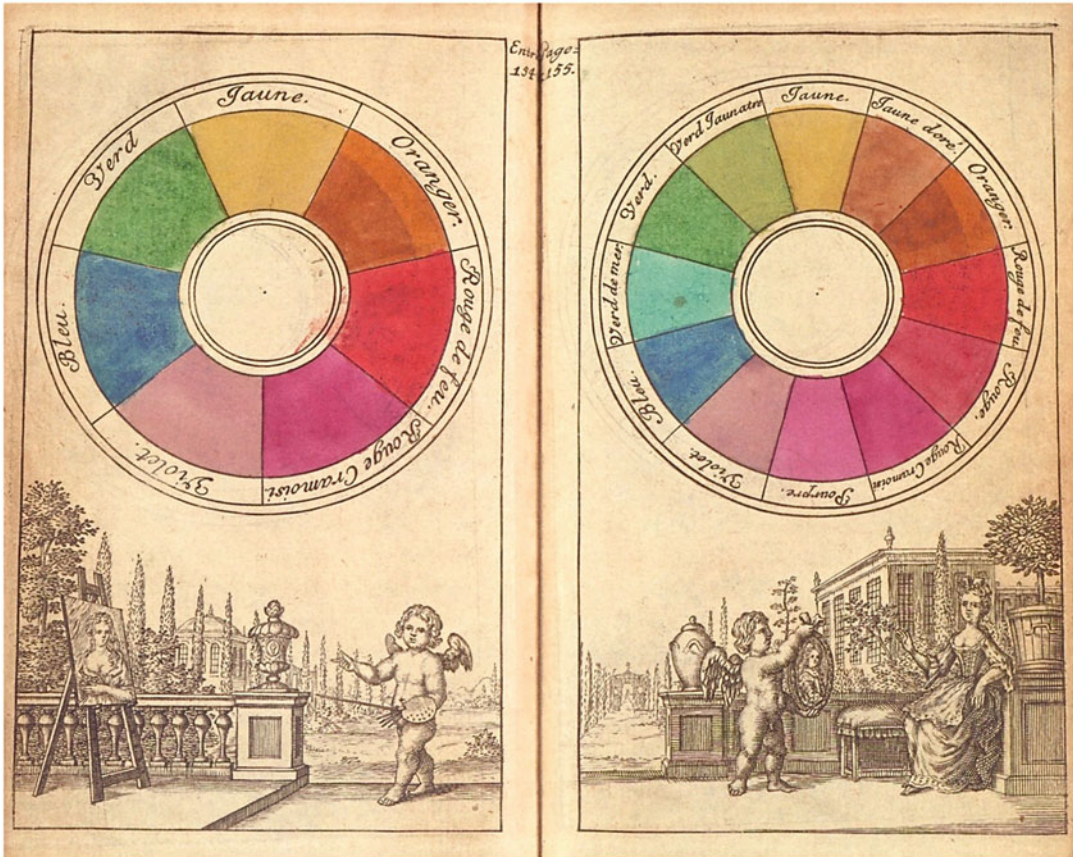
Color Circle,

Fig. 6 Michel-Eugène Chevreul 1864. Color circle with 72 discrete hues. (Reproduced from Ref. [6], with permission)



hues from just three “primitive colors” was already known at that time so the use of two “primitive” reds is striking. Perhaps the author did not think that any available red pigment could qualify as “true red” so two reds had to be

used, one a yellowish red and the other bluish. And perhaps no single red could deliver a satisfactory orange as well as a satisfactory purple. In this respect these early color circles foreshadow the Color Bias Wheel devised by Michael Wilcox



Color Circle, Fig. 7 Color circles from the enlarged 1708 edition of the *Treatise on Miniature Painting*. (Reproduced from Ref. [6], with permission)

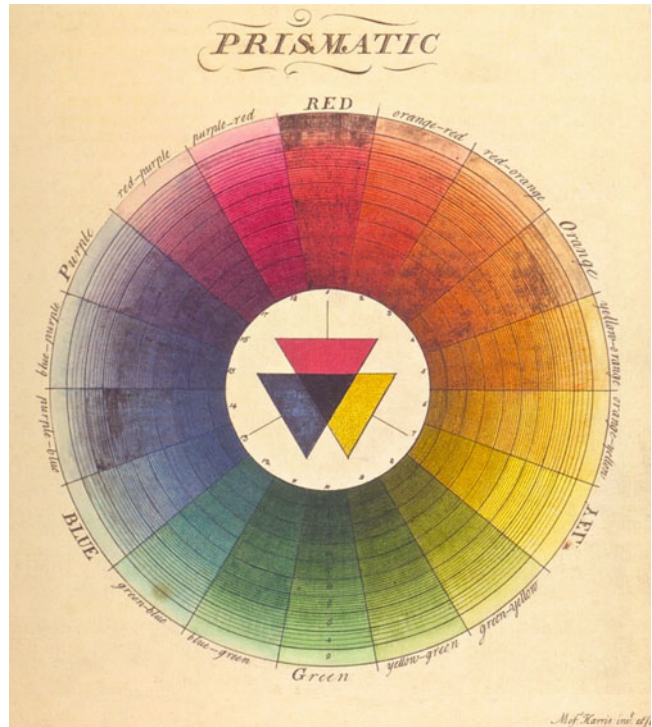
[10, p. 15]. To make sure that an artist can achieve vivid colors all round the circle, Wilcox has two blues and two yellows as well as two reds in his Color Bias Wheel.

Harris

Moses Harris held to the belief that “all the variety of colours . . . can be formed from Red, Blue, and Yellow” [11, p. 3], but he seems to have recognized the shortcomings of available pigments. He explains that he “treats on colour in the abstract” [11, p. 7]. This suggests that he had in mind some kind of theoretical ideal. As he points out, “Colour which we may call material, or artificial, are very imperfect in themselves, and being made of

various substances . . . maketh the colouring part extremely difficult . . .” [11, p. 7]. For Harris, red, blue, and yellow were “primitives” which could be mixed to produce the “mediates” purple, green, and orange. He does list representative pigments: vermilion, ultramarine, and king’s yellow for the “primitives” and sap green and red orpiment for two of the “mediates” (no pigment is listed for purple). The fact that he lists separate pigments for his “mediates” suggests that these pigments would have been used, in addition to vermilion, ultramarine, and king’s yellow, to paint the color circles in his book – this is to make his demonstration of the theory more convincing with acceptably vivid colors all round the circle (Fig. 8).

Color Circle, Fig. 8 Color circle by Moses Harris 1772. (Reproduced from Ref. [6], with permission)



Harris was an entomologist as well as an accomplished artist. Reference to his color circles would have been helpful when identifying and recording the colors of butterflies and other insects, while the circles could also serve as a guide to mixing paints to match those colors. Furthermore, Harris may have been the first to point out that a color circle can reveal relationships between colors that would now be called complementary. He refers to “contrasting colors” that are “so frequently necessary in painting” [11, p. 6] and goes on to explain how one should “look for the colour . . . in the system, and directly opposite to it you will find the contrast wanted” [11, p. 6]. And he provides a kind of definition: “if the colours so mixed are possest of all their powers, they then compose a deep black” [11, p. 7]. One current definition of complementary colors is that they should mix to a neutral – white from additive mixture in the case of lights, gray from partitive mixing in the case of colored segments on a spinning disc, and near black from subtractive mixing in the case of paints. At the end of the book, Harris

describes the phenomenon of colored shadows. He explains that a stick placed in the orange light of a candle will cast a blue shadow, this result being predictable from the positions of orange and blue on opposite sides of his circle. A more extensive discussion of complementary colors can be found under that heading in this encyclopedia.

Goethe

Johann Wolfgang von Goethe is best known as a writer of novels, plays, and poetry, but he also wrote a book on color theory which remains influential today. Goethe was satisfied that “yellow, blue, and red, may be assumed as pure elementary colours, already existing; from these, violet, orange, and green, are the simplest combined results” [12, p. 224]. When these six colors are arranged in a circle, yellow is opposite violet, blue is opposite orange, and red is opposite green. Goethe places great emphasis on such relationships: “the colours diametrically opposed to each

other in this diagram are those which reciprocally evoke each other in the eye” [12, p. 21]. This can be recognized in afterimages and Goethe describes his experience of a “beautiful sea-green” when a girl wearing a scarlet bodice moved out of sight [12, p. 22]. He argues that such experiences show how the eye demands completeness. The red bodice gave way to green as a union of blue and yellow. Goethe saw in afterimages “a natural phenomenon immediately applicable to aesthetic purposes” [12, p. 320]. The afterimage phenomenon provides another way of defining complementary relationships with the added claim by Goethe that such relationships are beautiful. So Goethe is reinforcing the notion that the color circle can be seen as a tool for developing harmonious color combinations.

Chevreul

Michel-Eugène Chevreul developed a set of nine color circles, graded from full hue (Fig. 6) to almost black, and each with 72 hues, as the basis of a comprehensive color order system. In his introduction to a translation of Chevreul’s book *De la loi du contraste simultané des couleurs*, Faber Birren explains how “Chevreul devoted himself not only to color organization, color harmony, and contrast effects, but to methods of naming and designation of colors” [13, p. 29]. A special memorial edition of Chevreul’s book was published by the National Press of France in 1889 when Chevreul himself was 103 years old. The potential confusion between the physical and visual aspects of color is evident in a prefatory note to this edition: “In order to guarantee to the plates of this book the stability which their scientific nature requires . . . it was necessary to resort only to mineral colors whose stability was certain. . . . Since the three colors chosen by Chevreul as basic, *red, yellow, blue*, cannot be reproduced precisely by means of isolated materials, they were obtained by mixing” [13, p. 27]. If the “basic” colors red, yellow, and blue could only be obtained by mixing, their status as “basic” would need to be clarified.

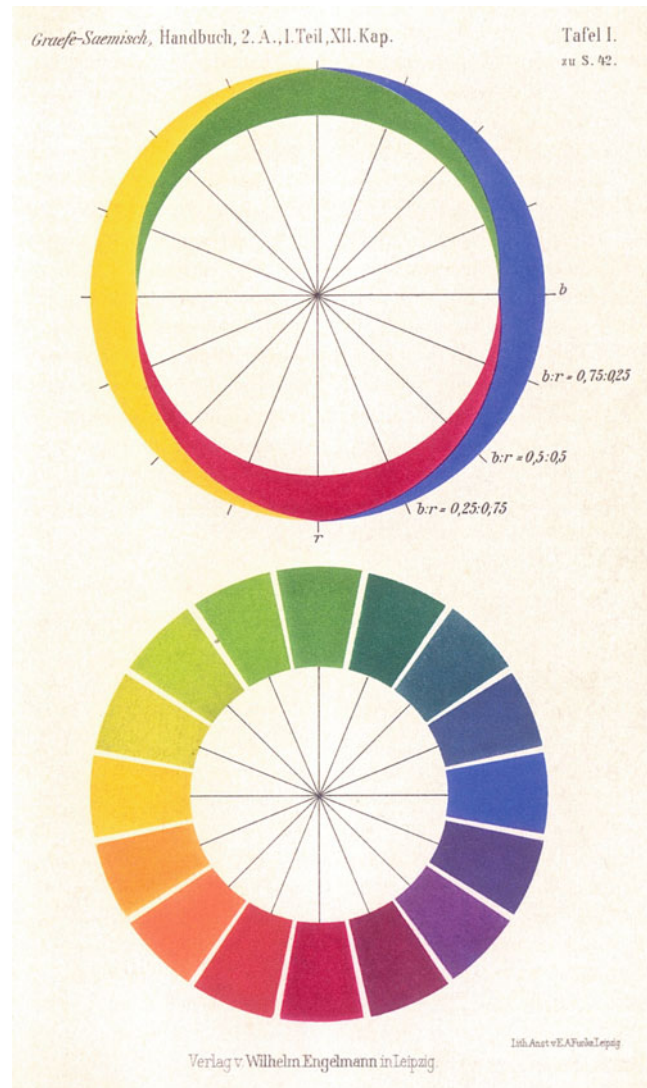
Hering

The primacy of red, yellow, and blue was challenged by Ewald Hering. The scientific orthodoxy that emerged during the nineteenth century was that three primary colors had their counterparts in the human eye in the form of three different types of receptor cell, each tuned to one of these primaries. This notion was first proposed by George Palmer when he suggested that “the surface of the retina is compounded of particles of three different kinds, analogous to the three rays of light” [14, p. 41]. Thomas Young, working independently, came to a similar conclusion. Young suggested that the sensitive particles in the retina were associated with “the three principal colours, red, yellow, and blue” [15, p. 147]. In a subsequent lecture he referred to “three simple sensations . . . red, green and violet” [16, p. 440]. Hering could not reconcile any set of three primary colors with his own subjective experience. For Hering there are six basic color phenomena – six *urfarben* – white, black, yellow, red, blue, and green. Since it is possible to describe the hue of any color in relation to yellow, red, blue, and green, Hering proposed that “corresponding to the four hue variables . . . there are four physiological variables” [17, p. 48]. Hering, therefore, designed a color circle which represented colors simply as visual phenomena with yellow, red, blue, and green as primaries (Fig. 9).

The Natural Color System (NCS)

Hering’s theories were developed in Sweden and are the basis for the Natural Color System (NCS) (Fig. 10). As with Hering’s color circle, the NCS has four “elementary colors” which are defined in visual terms as a yellow that is neither greenish nor reddish, a red that is neither yellowish nor bluish, a blue that is neither reddish nor greenish, and a green that is neither bluish nor yellowish [18, p. 132]. It is important to note that the NCS was designed just as a means of describing colors and showing how they are related as visual phenomena. The NCS is “value neutral in that it does not give rules for what is ugly and what is

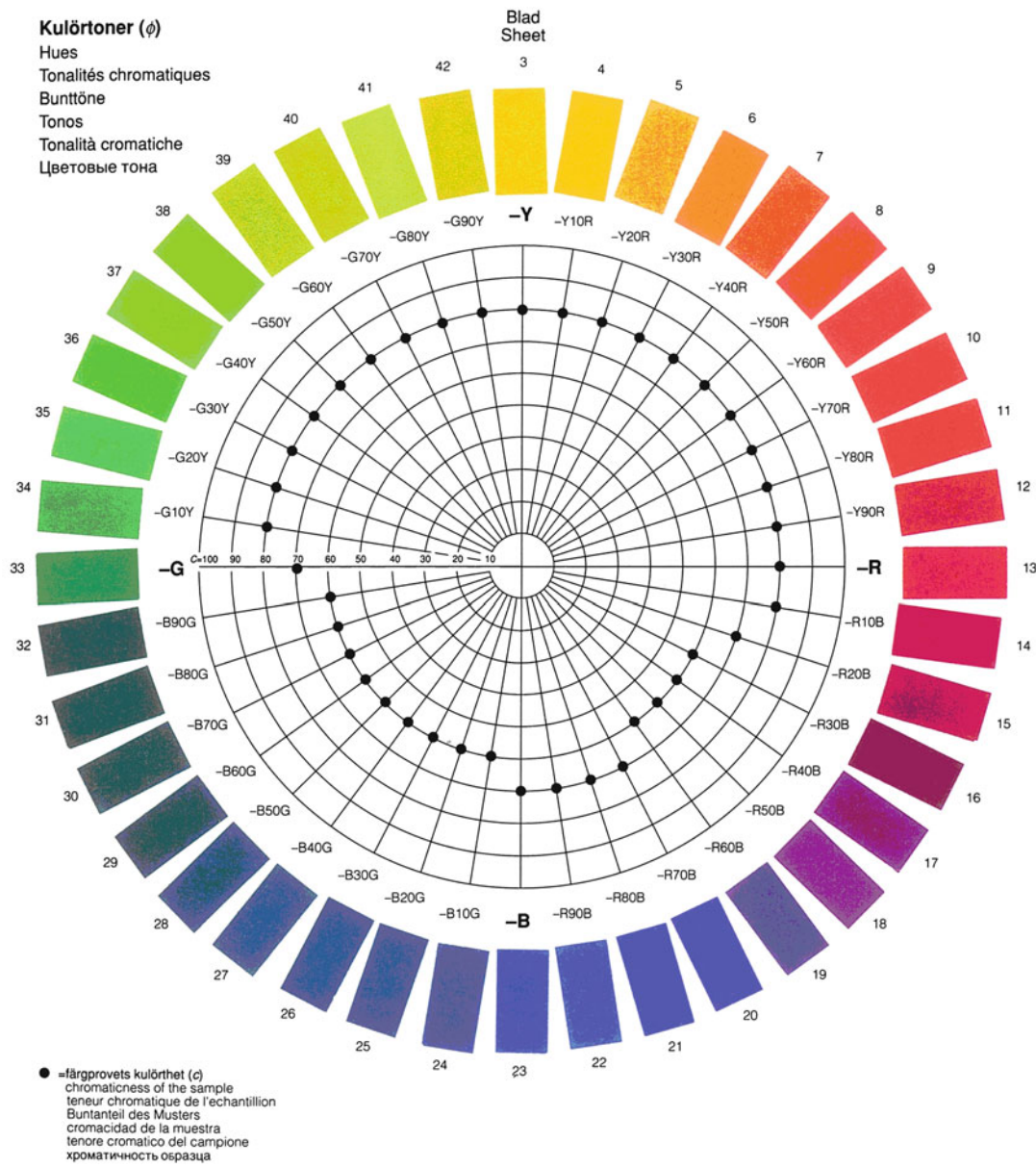
Color Circle, Fig. 9 Color circles by Ewald Hering. (Reproduced from Ref. [6], with permission)



attractive” [19, p. 4]. There are no claims that it is to be used as a guide to color harmony. It has been criticized for not having the degree of difference between neighboring colors the same all round the circle. Although the colors are shown in a continuous circle, the four elementary colors are to be understood as the beginnings and ends of four separate hue sequences. The colors between red and blue are equally spaced visually as are the colors between blue and green, but there is a greater degree of visual difference in the red to blue sequence than in the blue to green.

Ostwald

Wilhelm Ostwald developed a color circle that is superficially similar to that of Hering and the NCS in that it is based on four rather than three “fundamental colors” (Fig. 11). However, these colors are not treated as the beginning and end points of four separate hue sequences but simply as landmarks in one continuous sequence. For Ostwald, the color opposite to a given color should be the one that is “most different” so that “the entire circle is filled with such pairs of contrasting



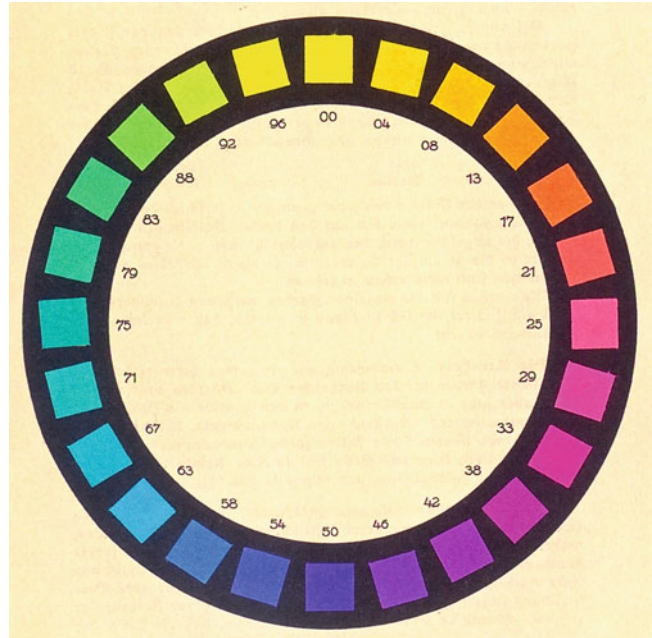
Color Circle, Fig. 10 Color circle of the Natural Color System (NCS). (Reproduced from Ref. [6], with permission)

colors, which shall be called complementary colors” [20, p. 34]. The complementary relationship is established by optical mixture using a spinning disc. Segments of yellow and red on a disc would blend to a single color when the disc is spun, the blend in this case appearing orange. The complementary relationship is established when

the blend appears a neutral gray. Ostwald chose “a pure yellow that is neither greenish nor reddish” [20, p. 33] as the starting point for his hue sequence. This would correspond to the elementary yellow of Hering and the NCS but the complementary of this yellow, as established by the spinning disc technique, is not quite a blue that is

Color Circle,

Fig. 11 Color circle by Wilhelm Ostwald.
(Reproduced from Ref. [6], with permission)



neither reddish nor greenish but one that is slightly reddish which Ostwald identifies as “ultramarine blue.” Similarly if Ostwald’s fundamental red is neither yellowish nor bluish, its complementary is what Ostwald calls “sea green,” a green that is certainly bluish. So Ostwald’s four fundamental colors are not the exact equivalent of the Hering/NCS elementary colors. And Ostwald, unlike those who developed the NCS, did intend his system to be used for generating harmonious color combinations for application in the arts and design. For Ostwald his system represents “order” and he is famous for his “basic law”: “Harmony = Order” [20, p. 65].

Müller

Ostwald’s ideas were taken up and developed by Aemilius Müller who produced his *Swiss Colour Atlas* using dyes rather than pigments. Müller’s color circle has 60 hue steps (Fig. 12). Müller also produced a number of designs with beautiful color gradations to demonstrate Ostwald’s “basic law.” Müller’s work is described by Stephanie Wettstein [21, pp. 144–149].

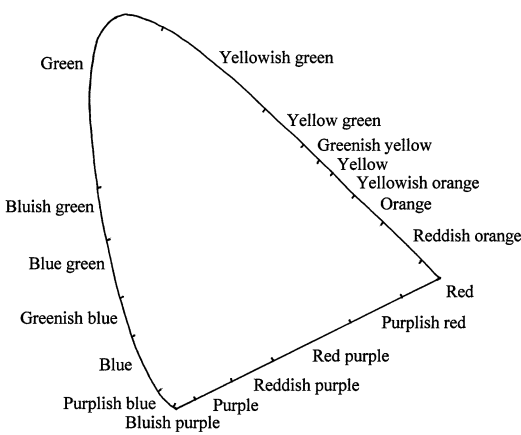
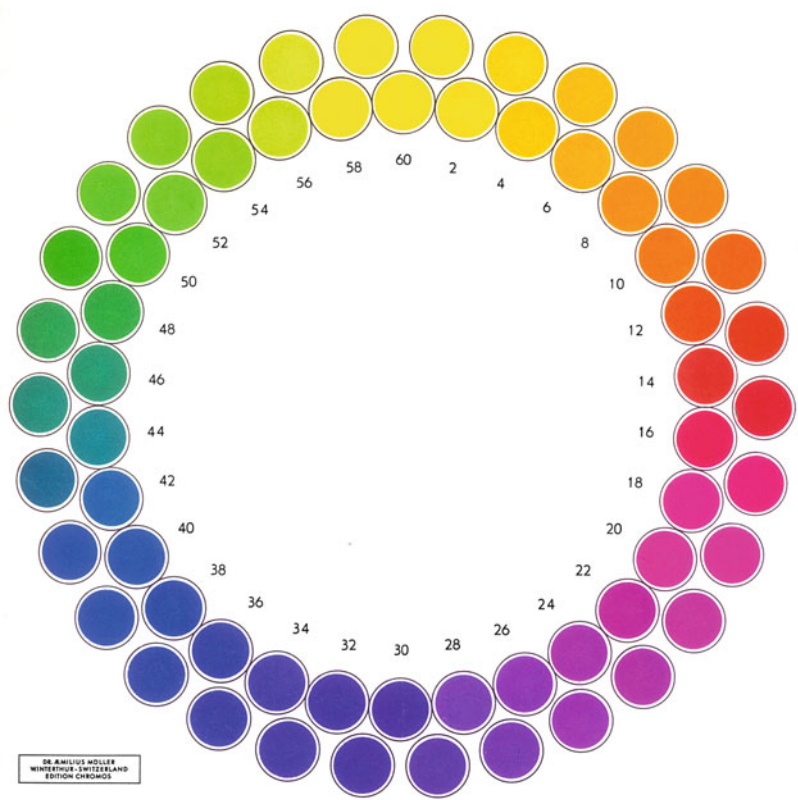
The CIE Chromaticity Diagram

Although not geometrically circular, the 1931 CIE chromaticity diagram can be regarded as a color circle in that it represents the pure spectral hues and the extra-spectral purples in a continuous sequence. The diagram is used in the international system for measuring color stimuli. The diagram is often shown in color but Roy Berns warns against this as being misleading [22, p. 61]. No printing inks can match the purity and intensity of the spectral lights themselves. It is better simply to show the line diagram marked out with color names, such as those proposed by Kenneth Kelly [23, p. 67], much as Newton did with the first color circle. The CIE diagram, with Kelly’s names, is shown in Fig. 13.

Munsell

Albert Henry Munsell developed his color system at the turn of the twentieth century and published *A Color Notation* in 1905. Munsell’s color circle has “five principal hues” [24, p. 20]. These are red, yellow, green, blue, and purple, which are spaced at equal intervals around the circle. In his

Color Circle,
Fig. 12 Color circle by
Aemilius Müller.
(Reproduced from Ref. [6],
with permission)



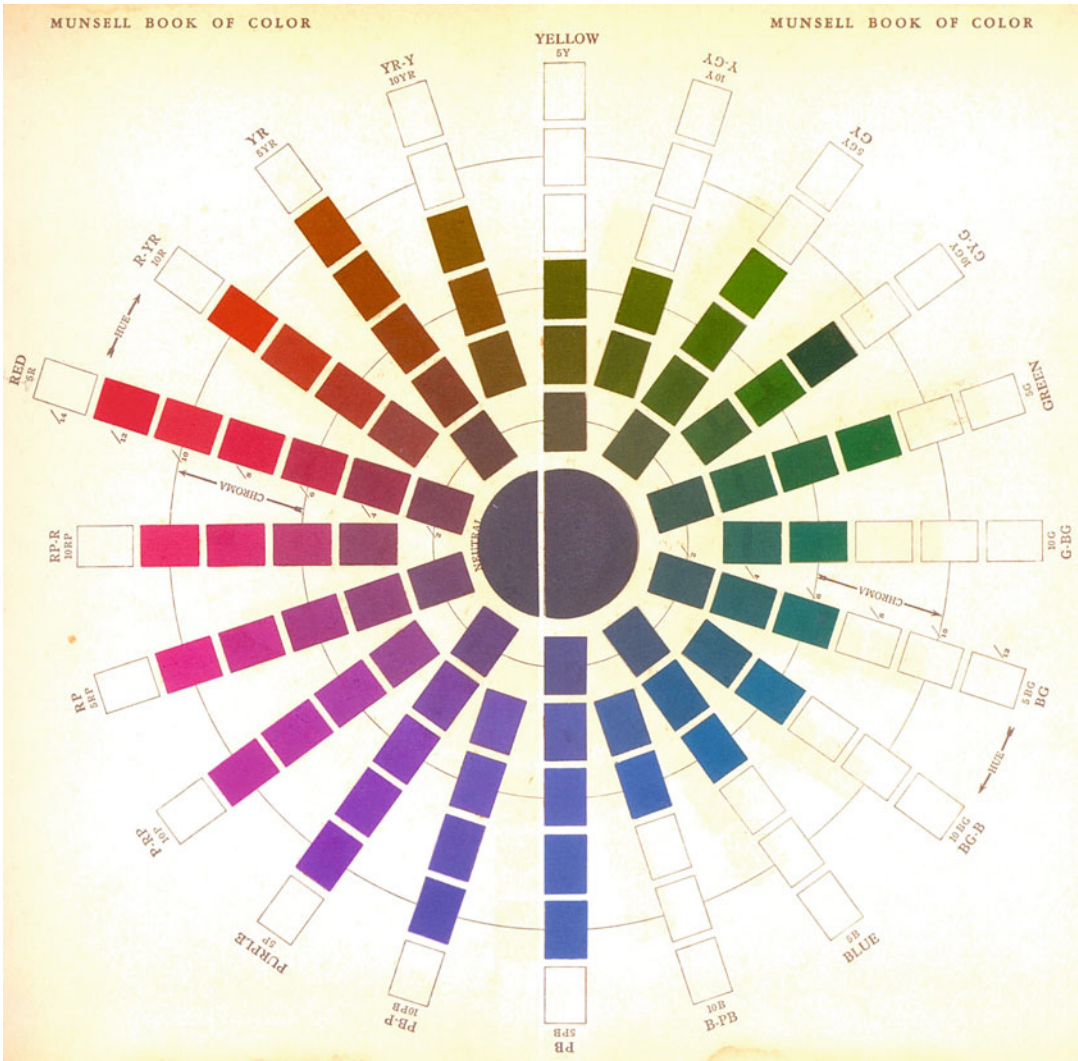
Color Circle, Fig. 13 CIE chromaticity diagram with
color names proposed by Kenneth Kelly

sequence of hues, Munsell aimed at perceptual uniformity [2, p. 115]. Like Ostwald, Munsell believed in an ordered arrangement of colors as the key to harmony and suggested a number of paths through his system that would connect

colors for a harmonious result. Figure 14 shows a page from the 1929 edition of the *Munsell Book of Color*. Twenty hues are included, each at several steps of increasing departure from neutral gray. Use of the CIE system to measure the Munsell color chips revealed a number of irregularities. A combination of instrumental measurement and visual judgments by members of an expert committee resulted in the Munsell renovations and the revised Munsell system that is widely used today.

The Color Circle Today

Today there are color circles in use that are based on three, four, and five primaries, but the three-primary model remains dominant with the primaries red, yellow, and blue; the secondaries orange, purple, and green; and 12 hues altogether. This is born out by an appeal to the Internet. Of the first 100 images from a Google search, made on



Color Circle, Fig. 14 Color circle from the 1929 edition of the *Munsell Book of Color*. (Reproduced from Ref. [6], with permission)

December 7, 2013, more than half were twelve-hue circles and nine of these were the circle designed by Johannes Itten with primaries and secondaries identified in the center (Fig. 15).

Itten

Itten's circle is attractive, clear, and memorable, but it needs to be viewed with caution. Itten was an artist writing for students of art. Artists work

with pigments and so the “color classification must be constructed in terms of the mixing of pigments” [25, p. 21]. But Itten defines his primaries in terms of appearance: “a red that is neither bluish nor yellowish; a yellow that is neither greenish nor reddish; and a blue that is neither greenish nor reddish” [25, p. 29]. Itten refers to green in these definitions, as does Hering in his definitions of *urfarben*, and green is one of the elementary colors of the NCS. Nevertheless, for



Color Circle, Fig. 15 Color circle by Johannes Itten as reconstructed by Lisa Hannaford

Itten, green is a secondary color, a mixture of blue and yellow.

From the experience of working with a number of different paints, it is possible to judge, from their appearance, how useful a group of paints will be in the mixing process. Students can be misled by the Itten diagram and may blame themselves if they are unable to mix a satisfactory range of colors from red, yellow, and blue primaries as these are defined by Itten. As Harald Arnkil points out, “anyone who has tried to create Itten’s twelve-hue color circle according to his requirements will be frustrated to a lesser or greater degree” [26, p. 88]. A more satisfactory range of colors can be mixed with paints that are closer in appearance to the cyan, magenta, and yellow inks as used by printers. This is demonstrated in the entry on ► [“Primary Colors”](#) in this encyclopedia.

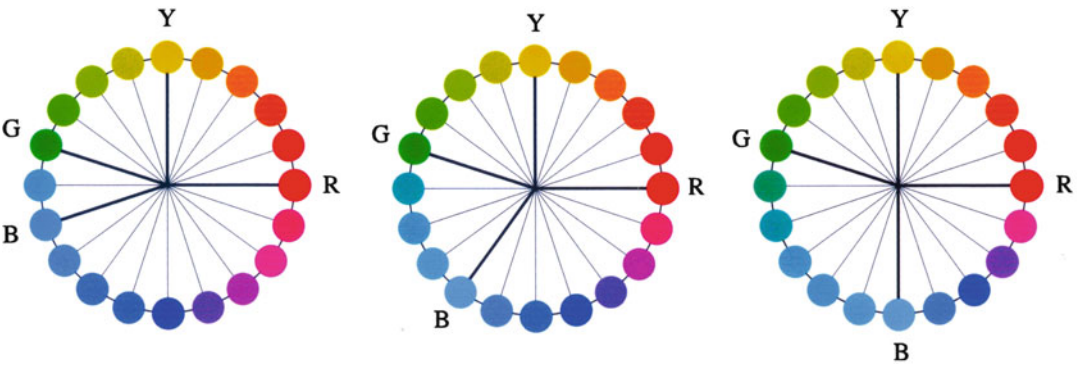
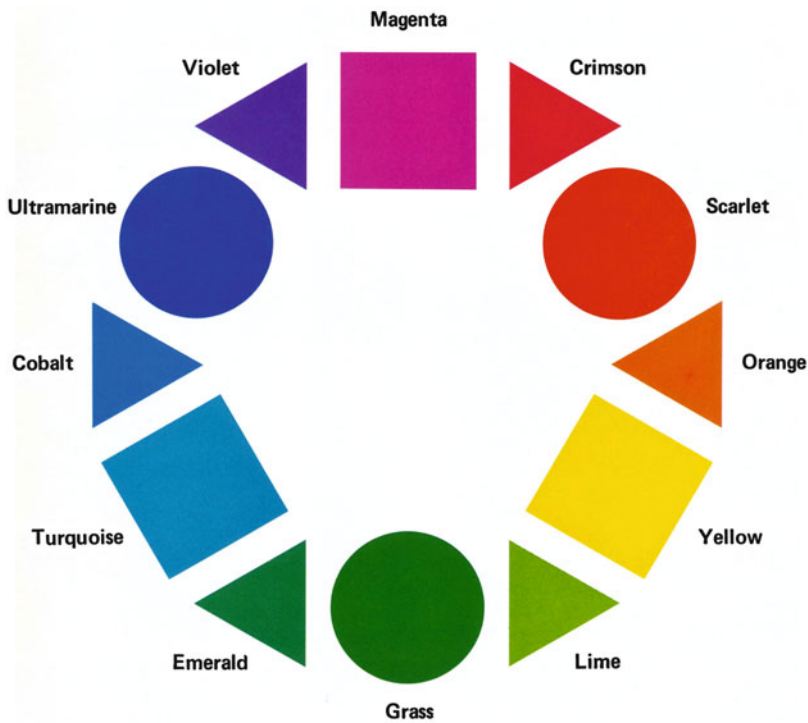
Itten’s color circle (Fig. 15) was recreated for the present entry by Lisa Hannaford using the computer program *Illustrator*. As displayed on the computer screen, Itten’s diagram is created with the additive primaries red, green, and blue. When printed in hard copy, for Itten’s book as well as from the computer file, the inks used are the subtractive primaries cyan, magenta, and yellow. The relationship between the additive primaries

and the subtractive primaries are shown in a color circle first presented at a conference in 1978 [27, p. 167, 28, p. 3 and cover] (Fig. 16). For this diagram, cyan is identified as turquoise, which is a more familiar color name. A shape code identifies the additive primaries as circles and the subtractive primaries as squares.

Perhaps Itten could have acknowledged the shortcomings of available pigments and followed the example of Moses Harris by explaining how he “treats on colour in the abstract.” If the diagram is misleading as a guide to mixing paints, it is reasonable to wonder what purpose it is intended to serve. Arnkil asks this question and suggests that it may have been “associated with his idea of the 12-colour circle as the basis of colour harmony” [26, p. 88].

Itten follows Goethe in asserting the significance of the afterimage phenomenon for color harmony, and he claims that opposite colors in his circle are complementary. Harmonious color combinations can supposedly be found by drawing a regular geometric figure inside the circle. Lines passing through the center of the circle, equilateral triangles, isosceles triangles, squares, and rectangles, as they touch the colors in the circle, all point to harmonious color combinations. As well as complementary pairs, Itten illustrates harmonious triads and tetrads – combinations of three and four colors. No doubt this is a useful starting point for people who lack confidence, but closer scrutiny reveals a problem. Itten defines complementary relationships in three ways, all of which have been encountered in the work of others: subtractive mixing to near black (Harris), afterimages (Goethe), and additive mixing to neutral gray (Ostwald). The theory would be more convincing if these different ways of defining complementary relationships yielded the same pairs, but this is not always the case. The most dramatic variation is with blue. The complementary of blue is red-orange from subtractive mixture, yellow-orange as an afterimage, and yellow from additive mixture. If these pairs are to be opposite to each other in different circles, the distribution of hues would need to be adjusted.

Color Circle,
Fig. 16 Color circle by
Paul Green-Armytage
showing the relationship
between additive and
subtractive primaries



Color Circle, Fig. 17 Color circles with complementary
colors opposite to each other according to three definitions.
From *left to right*: subtractive mixture to near *black*; the

generation of each other's hue in afterimages; optical mix-
ture to neutral *gray*

An Elastic Color Circle

Complementary relationships, as established in
different ways, can be illustrated by stretching
and compressing the color circle [29, p. 266].
With the elementary colors of the NCS as refer-
ence points, and by keeping yellow and red

constantly at 12:00 o'clock and 3:00 o'clock,
respectively, blue and green can be moved to
new positions. The number of hue steps between
the elementary colors will increase or decrease
accordingly to bring the differently defined com-
plementary pairs opposite to each other as shown
in Fig. 17.

Conclusion

The color circle has a long history and is well established in color theory. It is used to represent colors as something physical, as lights, pigments, inks, or dyes, as well as colors as visual phenomena. It is used to illustrate relationships between colors, in systems of color classification and identification, as a guide to mixing and as a tool in the search for harmonious color combinations. Although the sequence of hues is always the same, the intervals between hues can vary as the principles behind the construction of the circles varies. There is no single color circle that is “correct.” Rather than try to establish a single color circle as some kind of standard, or to insist on a single purpose for the color circle, it is more helpful to recognize that a color circle can embody a variety of information and that it may be necessary to stretch or compress the hues in the circle according to the information that is required.

Cross-References

- [Appearance](#)
- [Chevreul, Michel-Eugène](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Complementary Colors](#)
- [Hering, Karl Ewald Konstantin](#)
- [Itten, Johannes](#)
- [Munsell, Albert Henry](#)
- [Ostwald, Friedrich Wilhelm](#)
- [Palmer, George](#)
- [Primary Colors](#)
- [Unique hues](#)

References

1. Newton, I.: Opticks. In: MacAdam, D. (ed.) Sources of Color Science. The MIT Press, Cambridge, MA (1970 [1704])
2. Kuehni, R., Schwarz, A.: Color ordered. Oxford University Press, New York (2008)

3. Russell, B.: History of Western Philosophy. George Allen & Unwin, London (1961)
4. Gage, J.: Colour and Culture. Thames and Hudson, London (1993)
5. Hård, A.: Quality Attributes of Colour Perception. Fackskrift Nr 8. The Swedish Color Centre Foundation, Stockholm (1969)
6. Spillmann, W. (ed.): Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann. Schwabe Verlag, Basel (2009)
7. Parkhurst, C., Feller, R.L.: Who invented the color wheel? Color. Res. Appl. 7(3), 217–230 (1982)
8. Spillmann, W.: Farbskalen – Farbkreise – Farbsysteme. Schweizerischer Maler- und Gipserunternehmer-Verband, Wallisellen (2001)
9. Schindler, V.M.: Michel Eugène Chevreul. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
10. Wilcox, M.: Blue and Yellow don’t Make Green. Artways, Perth (1987)
11. Harris, M.: The Natural System of Colours. Whitney Library of Design, New York (1963 [1766])
12. Goethe, J.W.: Theory of Colours. The MIT Press, Cambridge, MA (1970 [1840])
13. Birren, F.: Introduction. In: Chevreul, M.E. (ed.) The Principles of Harmony and Contrast of Colors and their Applications to the Arts. Reinhold Publishing Corporation, New York (1967 [1854])
14. Palmer, G.: Theory of colors and vision. In: MacAdam, D. (ed.) Sources of Color Science. The MIT Press, Cambridge, MA (1970 [1777])
15. Young, T.: Theory of light and colours. In: Peacock, G. (ed.) Miscellaneous Works of the Late Thomas Young. John Murray, London (1855 [1801])
16. Young, T.: A Course of Lectures on Natural Philosophy and the Mechanical Arts. Johnson Reprint Corporation, New York (1971 [1807])
17. Hering, E.: Outlines of a Theory of the Light Sense. Harvard University Press, Cambridge, MA (1964 [1920])
18. Hård, A., Sivik, L.: NCS – natural color system: a Swedish standard for color notation. Color. Res. Appl. 6(3), 129–138 (1981)
19. Smedal, G.: NCS – as a basis for colour training at the National College of art, craft and design, Bergen, Norway. In: Hård, A., et al. (eds.) Colour report F 26. Scandinavian Colour Institute, Stockholm (1983)
20. Ostwald, W.: The Color Primer. Van Nostrand Reinhold, New York (1969)
21. Wettstein, S.: Aemilius Müller. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
22. Berns, R.: Billmeyer and Saltzman’s Principles of Color Technology. Wiley, New York (2000)
23. Agoston, G.A.: Color Theory and its Application in Art and Design. Springer, Berlin (1987)
24. Munsell, A.H.: A color notation. Munsell Color, Newburgh (1946)

25. Itten, J.: *The Elements of Color*. Van Nostrand Reinhold, New York (1970)
26. Arnkil, H.: *Colours in the Visual World*. Aalto University, Helsinki (2013)
27. Green-Armytage, P.: Violets aren't blue: colour sensations and colour names. In: Condous, J., et al. (eds.) *Arts in Cultural Diversity*. Holt, Rinehart and Winston, Sydney (1980)
28. Green-Armytage, P.: Violets aren't blue, they're 'purple. *Gaz. Off. J. West. Aust. Inst. Technol.* **12**(1), 2–6 (1979)
29. Green-Armytage, P.: The value of knowledge for colour design. *Color. Res. Appl.* **31**(4), 253–269 (2006)

Color Classification

► Color Constancy and Categories

Color Combination

José Luis Caivano
 Universidad de Buenos Aires, and Conicet,
 Secretaría de Investigaciones FADU-UBA,
 Buenos Aires, Argentina

Synonyms

Color coordination; Color harmony; Color mixture; Color syntax; Color union

Definition

Color combination is mainly an aspect of color syntax. To combine means to put one thing in relation to another, or several things together, so that the individuals lose significance and the meaning of the whole predominates. To combine also means to organize an ordered sequence. In some cases, it acquires the sense of mixing or merging. However, mixing color pigments or lights normally yields just one color as the result of the mixture, and in this sense, it cannot be termed a color combination, for which two or more colors in some relation must be perceived.

Combination certainly is very closely connected to harmonization and coordination. Color combination, thus, is meant whenever there is more than one color associated, related, or harmonized with another: Two colors already determine a certain kind of combination.

Overview

Color combination is, in principle, an aspect of color syntax. All perceptible colors can be organized in the so-called **color order systems** or models. This is usually made by means of three color variables or dimensions, for instance, hue, saturation, and lightness (HSL), or hue, blackness, and chromaticness (according to the Natural Color System), or hue, value, and chroma (according to the Munsell color system), or some other similar triad of variables. These color order systems allow for the precise identification and notation of colors and their arrangement in a logical way. Some of the systems even allow predicting the results of color mixtures. It is possible to compare color order systems to dictionaries, which assemble all the words available in a language in alphabetical order. Therefore, color order systems arrange and organize all the colors that humans can see, produce, and use, according to certain sequences determined by the mentioned color variables.

This possibility of having the repertoire of all perceptible colors orderly arranged facilitates the selection of colors, by following certain criteria, in order to use them in artistic compositions, architectural works, or pieces of design. Only in few rare occasions (for instance, in experimental situations) colors are seen isolated; in the great majority of cases (both in nature and in human productions), colors appear in a context where there are also other colors. That is to say, every color is combined in a certain way with other colors. Such as words (which in a dictionary appear isolated) are combined with other words in order to make sentences with some sense and give origin to poetry, narrative, essay, etc., and in the same way as sounds are combined by a music composer to give origin to musical pieces, so

colors are grouped in larger compositional units. It is the context, the particular combination, the way in which colors are grouped together and related to one another, what gives a certain kind of signification or meaning, some utility to the whole composition and to each of the involved colors.

Painters dispose and mix colors in their ► **palette** with the final aim of combining them in a canvas. Architects combine materials with different natural colors in a building or use paints to endorse different parts of their work with color. Filmmakers and directors of photography decide about the color sequences that appear in successive scenes of the film. Clothing and fashion designers think about the chromatic combinations of the apparels they produce. Landscape designers choose and arrange the botanical species and other materials taking into account also color combinations. It is possible to continue providing this kind of examples almost indefinitely, because there is practically no profession, discipline, or human activity in which color does not play a role.

If according to different authors and experiments, the number of perceptible colors may range, depending on various factors, from 2,000,000 to 7,000,000, the combinatorial possibilities rise to billions, even restricting them to the minimal expression of just two- or three-color combinations. Now, how are colors combined, with which kind of criteria, and in which type of contexts? At first, it is possible to talk about spatial and temporal contexts, depending on the colors being arranged simultaneously in an object or successively in a certain temporal sequence.

Syntactic Color Combinations

Spatial Color Combinations

The spatial color combinations have, at first, some basic and elementary rules. In terms of the abstract and logical possibilities and from the point of view of spatial arrangements, three possible cases can be pointed out for two-color combinations in a two-dimensional space:

1. That one color is applied over and inside another (interiority).

2. That both colors partially overlap each other (overlapping).
3. That they are juxtaposed one beside the other (juxtaposition).

The possibility of both colors being some distance apart is not considered here because in this case the color filling the separation, the background, appears as a third color. In addition, there is no sense in considering a total superposition of both colors (both occupying exactly the same space), because in this case the result is just one color, and hence, this cannot be termed a color combination.

These three possibilities produce different results or have different consequences for color light and for color-pigment combination and also differ if there is a mixture or blending of the involved colors or if opaque color surfaces that do not mix together are combined. Combining colors imply that in some cases the colors are mixed and give origin to new colors. However, if the result of the mixture is just one color, this will not be a color combination.

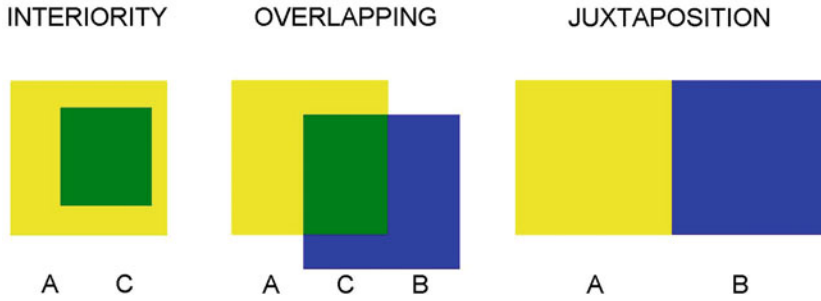
For instance, considering transparent color filters:

1. If over an area of a transparent color filter *A* another piece of color filter *B* is set in relation of interiority, the outcome is two colors: color *A* and a new color, *C*, which is the subtractive mixture of *A* and *B*, while color *B* is missed.
2. If the colors overlap, the result is three colors: *A*, *B*, and *C*.
3. In the case of juxtaposition, there is no color mixture, so that the result is color *A* just besides color *B*.

These combinations with their respective results are shown in Fig. 1.

The same happens with transparent inks and watercolors. Similar situations arise also when *A* and *B* are colored lights, but in this case color *C* is the result of an additive mixture. The situations are quite different with opaque color surfaces: In all cases, no new color *C* appears.

Additionally, in all situations – interiority, overlapping, and juxtaposition – phenomena of



Color Combination, Fig. 1 Basic possibilities of two-color combinations in a two-dimensional space

simultaneous contrast occur. Thus, considering colors *A* and *B* in isolation, the perceptual result of their combination is colors *A1* and *B1* (apart from the cases in which color *C* appears), because each color is tinged with the complementary color of the other, according to the principles of simultaneous contrast. That is to say, color sensations change from seeing color stimuli in isolation to seeing color stimuli in combination.

Now, if the combined colors have relatively small areas to be perceived individually, three situations may arise: (1) An additive mixture is produced when they consist of small pixels or cells of color light (as in the case of the TV screen). (2) A partitive mixture occurs when they are small dots of surface colors (as with the pointillist technique of painting). And (3) a mixed syntheses occur – partially subtractive and partitive – if the small dots are made of transparent inks that in some zones overlap each other and in some others are separated on the background (as in the case of color printing).

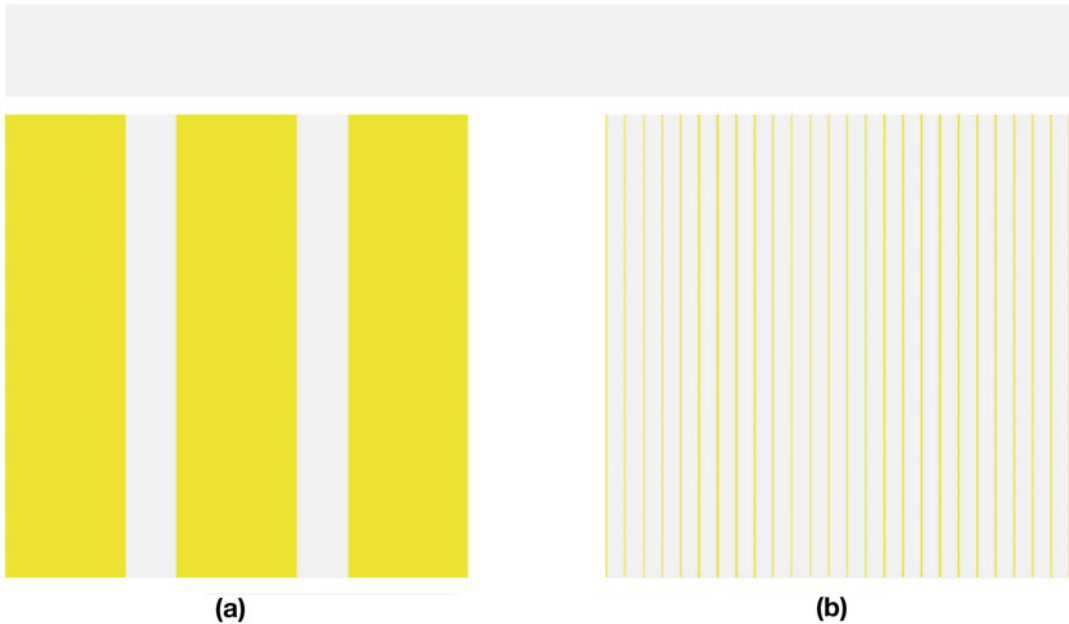
The variation in relative size of two color fields may produce two opposite perceptual phenomena: color assimilation (or color spreading effect) and simultaneous color contrast. Let us have an alternating sequence of two color strips, for instance yellow and gray, with different relative widths. When the yellow strip is wider than the gray one, a simultaneous contrast occurs, and the gray stripes are perceived as tinged with the complementary color of yellow, and look bluish (i.e., more dissimilar to yellow). When the yellow stripes are very thin lines on the gray background an assimilation is produced, and the same gray

looks now yellowish (i.e., more similar to yellow), when viewed from an appropriate distance (Fig. 2). Actually, in Fig. 2b the yellow thin lines start to blend with the gray background or spread over it because we are approximating a partitive or optical mixture. Color contrast increases the difference between two colors, while color assimilation increases the resemblance between them, and the appearance of one phenomenon or the other depends on the relative size of two combined colors.

Another aspect concerning the relative size of combined color fields is the role that certain colors play in a design. In three-color combinations, the following categories are usually distinguished: dominant or main color, secondary or subordinate color, and accent color (Fig. 3). In architectural facades, interior decoration, clothing, and other branches of design, an empirical rule is often applied: approximately 60% of the surface is given to the dominant color, around 30% is left for the secondary color, while the accent color takes the remaining percent of the total surface (around 10%). In general, high saturation or chromaticness is used for the accent color, while the dominant and secondary colors are usually less chromatic. In addition, a complementary contrast of hue is often used between the accent color and the dominant and secondary pair. The key is to create a contrast in terms of lightness, saturation, and hue.

Temporal Color Combinations

Phenomena of contrast appear whenever two or more colors are combined in a certain relation, but



Color Combination, Fig. 2 Different proportions in the combination of two color stimuli may produce simultaneous contrast (left, where the gray looks bluish) or

assimilation (right, where the gray looks yellowish). Note: Both gray stimuli are equal to the horizontal stripe above

if this is a temporal combination, where the colors appear in a sequence, with certain durations and intervals, what is produced is a successive contrast or the phenomenon of post-image.

When the time span of visualization of a color followed by another is long enough, an adaptation to the first stimulus occurs, and, therefore, the second stimulus is affected by the successive color contrast.

When the frequency of appearance of two or more colors is fast enough to fall below the perceptive threshold (as in a flickering situation), an optical mixture of the colors involved in the sequence will be produced. It also happens here that two or more colors combined in these conditions give only one color as a result, the color that is the consequence of the optical mixture.

Color Selection as the First Step for Harmonic Color Combinations

If some specific chromatic relations among the combined colors are taken into account, the field of [color harmony](#) appears. There are a lot of proposals and theories about this. From a purely

syntactical point of view, paying attention to the relations among the colors themselves and the quantity of colors combined (two or more colors), it is possible to mention, for instance, a combination of monochromatic colors, complementary colors, split complementaries, double complementaries, analogs, color triads or trichrome combinations, tetrachrome combinations, etc.

César Jannello proposed criteria to select colors to be applied in a combination. Starting from the three color variables or dimensions – for instance, hue, saturation, and lightness – there are just eight possibilities for the selection of colors, whether these variables are kept constant or change. In Fig. 4, the sign plus (+) means constancy, and the sign minus (–) represents variation of the considered dimension. The first formula, the one in which everything is constant, is not of much use because it gives as a result the selection of just one and the same color. In the remaining formulas, where some type of variation appears, the interval of variation may be constant or may change according to some criterion, for instance, by modifying hue, lightness, or



Color Combination, Fig. 3 (a) Three-color combination in dress: blue 66% (dominant color), cyan 26% (secondary color), red 8% (accent color). (b) A facade with a three-

color combination: yellow 60% (dominant color), white 30% (secondary color), red 10% (accent color)

saturation in regular steps or intervals, by increasing intervals, by choosing opposite poles, etc. This model provides a logical basis to select and combine colors for design purposes.

A Theory for Colors in Combination

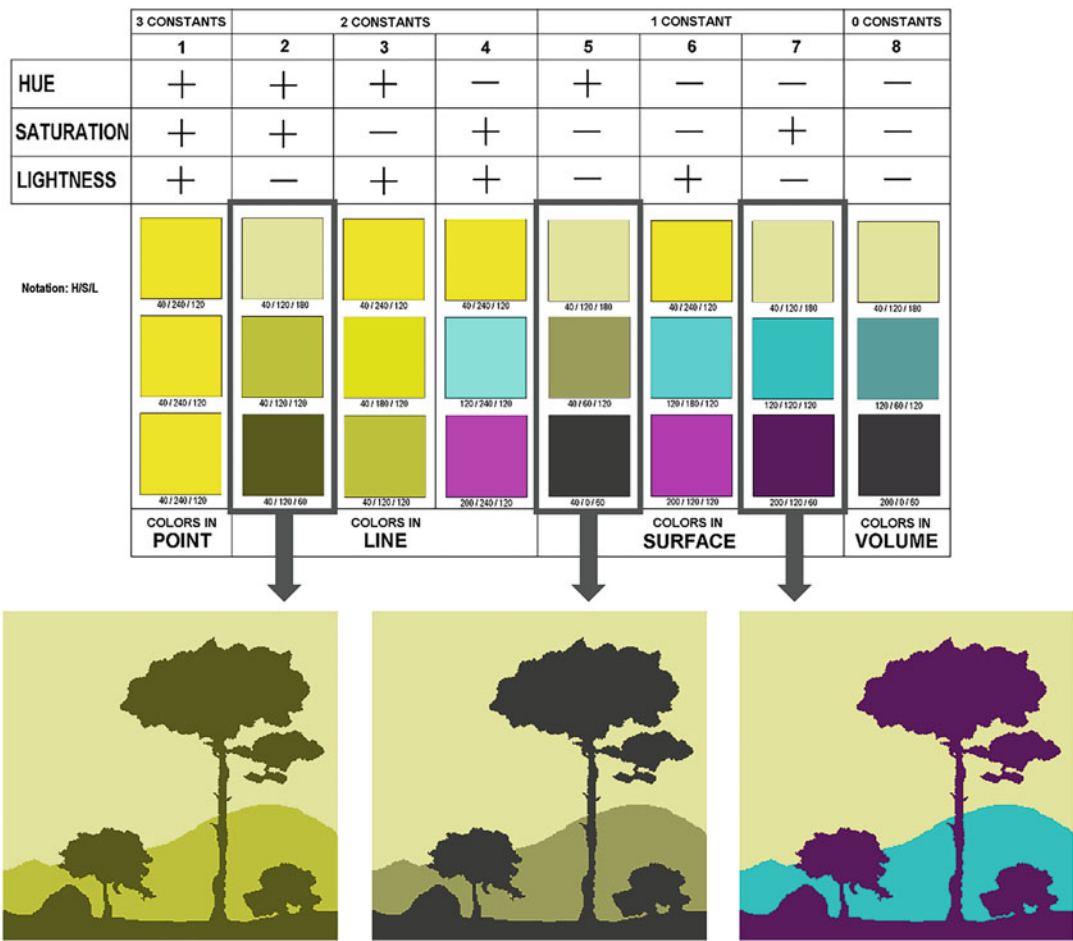
Anders Hård and Lars Sivik [1] have settled the basis for a theory of colors in combination. They have developed a structure that considers three dimensions or factors that are useful to describe or analyze color combinations: (1) *color interval* (dealing with color discrimination and having distinctness of border, interval kind, and interval size as subvariables); (2) *color chord* (dealing with color identification and having complexity, chord category, and chord type as subvariables); (3) *color tuning* (dealing with how color combinations can be varied and having surface relations, color relations, and order rhythm as subvariables).

Figure 5 shows an outline of this model, published by Hård in 1997 [2].

This model for color combination was worked out along various years, and during its development their concepts, dimensions, and subvariables were changing to some extent. Hård and Sivik published previous formulations of this theory in 1985 and 1994 [3, 4]. In some of these, for instance, the visual context in which the color combination appears is considered as a fourth and very important factor.

Semantic and Pragmatic Aspects in Color Combinations

At the beginning of this article, we have said that the combination of colors is mainly a syntactical aspect. But since colors have a semantic weight,

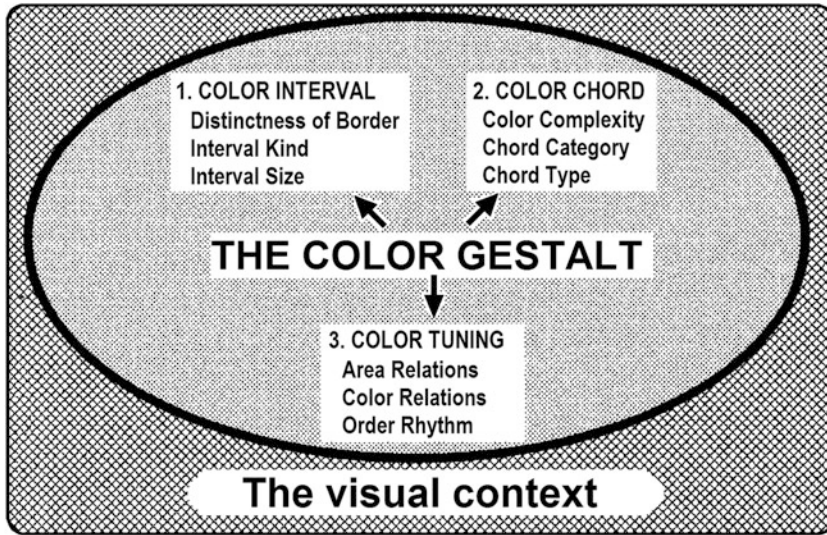


Color Combination, Fig. 4 One example of Jannello’s criteria for selecting colors to be used in a color combination

produce emotions, have meanings, are used as signals, indicate situations, propose behaviors, communicate messages, etc., and all this can vary according to the way in which colors are combined and according to the context, it is also possible to consider color combinations from a semantic point of view.

Here is a simple example. The three colors of the traffic lights (red, yellow, green) are a spatial and temporal syntactical combination, on one side. They make a triad of separated color lights, displayed in a circular shape, that appear with a spatially codified vertical arrangement: red at top, yellow in the middle, and green below. The temporal sequence is also regulated and codified: yellow comes after green and red comes after

yellow. The step from red back to green is normally direct, without intermediation of yellow, and the same sequence appears again: green, yellow, and red. The size of each color has also a specific relationship: the red circular light has a larger size than yellow and green. The same happens with the time interval or duration in which each light is displayed: yellow appears for a brief instant, while red and green have longer durations. Now on, all these are purely syntactic aspects. Nothing has been said yet about what meaning this selection, arrangement, and sequence of combined colors conveys. By entering in the semantic domain, it is possible to talk about the codification of those three colors in that particular context of use. Red means “stop,” green means “go,” while



Color Combination, Fig. 5 Outline of a color combination model by Anders Hård [2]

yellow is a warning about the change of light from green to red that is coming soon and imply that the user has to take the necessary caution, either apply the brakes to stop the vehicle or speed up the march to make it through before the red light appears.

Thus, there are also semantic issues that are combined from a syntactical arrangement of colors and the context in which they are used and interpreted. The same red color used for the traffic lights may have very different meanings in other contexts. It may connote emotions such as passion, love, and rage. It may indicate something that is important to notice and deserves to stand out (a red typo in a context of black letters and words). It means expulsion from the field in the context of a football (soccer) match. It may also connote speed or status in a car (a red Ferrari), etc. Hence, it is the context (either the social, cultural, geographic, or temporal context in which the colors appear, as well as the relationship with other colors that are in the same context or situation) what endorses colors and color combinations with a certain sense or meaning.

Color combinations have been studied from the semantic point of view by various authors. Elda Cerrato points out the basic concepts behind the idea of color combination, discusses some

color order systems and color harmony theories related to this (mainly Ostwald, Munsell, and Arnheim), and addresses the issue of how culture conditions harmonies, preferences, and aesthetical principles of color combinations [5].

Shigenobu Kobayashi, working at the Nippon Color and Design Research Institute, has devised a method to classify single colors or three-color combinations by their associated images [6, 7]. Through the analysis using the axes warm/cool, soft/hard, and clear/grayish as coordinates, this method, which was also extended and developed with some collaborators [8, 9], can plot climatic and cultural differences in color semantics. In 1997, Kobayashi and Iwamatsu extended the color combination research to be able to include five-color combinations [10], even when from the countless number of possible five-color combinations they chose to make the survey by selecting 20 pairs of contrasting combinations of five colors (40 color combinations in total). To justify that particular research, they point out that “a five-color combination makes it easier to gain a psychological understanding of scenic conditions and convey the sense of ambience than two- and three-color combinations.”

The book by Hideaki Chijiiwa, from the Musashino College of Art, at Musashino Art

University, intends to be a manual for choosing color combinations for different purposes, taking into account meanings and mood [11]. For practical applications in art, design, industry, or everyday life, it provides a guide for selecting two-, three-, and four-color combinations associated to adjectives such as striking, tranquil, exciting, natural, warm, cold, young, feminine, and surprising. As for the quantity of colors to be employed in a combination, the author advises to limit them to two or three. A warning is made when using four-color combinations that should be selected very carefully, while five-color combinations are directly discouraged. The book by Bride M. Whelan continues in the same venue [12].

Lars Sivik (working sometimes in collaboration with Anders Hård and Charles Taft) carried out research on the meanings of color combinations [13–15]. The descriptive model uses the Natural Color System as a basis, and the methods are aimed at studying the stability and variability of color-meaning associations across time and cultures. These studies “literally mapped the world of color with respect to how associations to various words systematically vary across different parts of the color world.” In the research published in 1989, Sivik selected 130 words by a semantic differential scaling method, and the subjects judged color images as “to how well the different word went with the color composition in question” [15]. The main purpose was “to obtain a small number of variables that would be reasonably representative of all color describing variables.”

Other authors have also used the semantic differential method to study the meanings of color combinations, and more specifically their affective values, by applying this tool to two- and three-color harmonies [16, 17].

Colors, in general, and color combinations, in particular, can be considered as a system of signs. They certainly have syntactic aspects (which include both the elements and the combinatory), semantic aspects, and pragmatic aspects that imply the use of these signs by the interpreters. In previous publications, the author of this entry describes and illustrates various semiotic concepts with examples taken from color theory and

provides an account of some of the advances of color theory within the framework of semiotic categories [18, 19].

Some color theorists go even beyond these considerations, proposing that colors and color combinations can be taken as a language, for instance, Luckiesh in 1918 [20], Sanz in 1985 and 2009 [21, 22], Oberascher in 1993 [23], as well as Hård in the already mentioned article published in 1997 [2]. However, this should be taken with certain caution. Human languages (for instance, verbal languages, English, Spanish, French, German, etc.) serve not only for communicational purposes but also for cognitive and modeling purposes; they allow to build categories, models, and theories about the world, in order to understand it, explain it, and make it meaningful for the human species. If it is possible to demonstrate that color combinations can have a similar status, then the color-language idea will be more than just a metaphor.

Cross-References

- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Color Scheme](#)
- [Complementary Colors](#)
- [Palette](#)

References

1. Hård, A., Sivik, L.: A theory of colors in combination – a descriptive model related to the NCS color-order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
2. Hård, A.: Colour as a language. Thoughts about communicating characteristics of colours. In: Sivik, L. (ed.) *Colour and Psychology*. From AIC Interim Meeting 96 in Gothenburg, pp. 6–12. Scandinavian Colour Institute, Stockholm (1997)
3. Hård, A.: A colour combination theory for a human environment. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC*. Centre Français de la Couleur, Paris (1985)
4. Sivik, L., Hård, A.: Some reflections on studying colour combinations. *Color. Res. Appl.* **19**(4), 286–295 (1994)

5. Cerrato, E.: Cultura y combinatorias de color: cómo la cultura condiciona armonías, preferencias, recomendaciones, leyes estéticas en las combinatorias del color. In: Caivano, J., Amuchástegui, R., López, M. (comp.) *Argen Color 2002, Actas del 6° Congreso Argentino del Color*, pp. 27–34. Grupo Argentino del Color, Buenos Aires (2004)
6. Kobayashi, S. (comp.): *A Book of Colors*. Kodansha international, Tokyo (1987)
7. Kobayashi, S.: *Color Image Scale*. Kodansha International, Tokyo (1991)
8. Kobayashi, S., Sato, K.: The theory of the color image scale and its application. In: Billmeyer, F., Wyszecki, G. (eds.) *AIC Color 77, Proceedings of the 3rd Congress*, pp. 382–383. Adam Hilger, Bristol (1978). <https://aic-color.org/resources/Documents/aic1977proc-reduced.pdf>. Accessed 24 Mar 2020
9. Kobayashi, S., Suzuki, H., Horiguchi, S., Iwamatchu, K.: Classifying 3-color combinations by their associated images on the warm/cool and clear/grayish axes. In: Nemesics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress*, vol. C, pp. 32–36. Hungarian National Color Committee, Budapest (1993)
10. Kobayashi, S., Iwamatsu, K.: Development of six methods of color psychological study. In: *AIC Color 97, Proceedings of the 8th Congress*, pp. 727–730. The Color Science Association of Japan, Kyoto (1997). <https://aic-color.org/resources/Documents/aic1997proc2-reduced.pdf>. Accessed 24 Mar 2020
11. Chijiwa, H.: *Color Harmony. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1987)
12. Whelan, B.M.: *Color Harmony 2. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1994)
13. Sivik, L.: Evaluation of colour combinations. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC*. Centre Français de la Couleur, Paris (1985)
14. Sivik, L.: Dimensions of meaning associated to color combination. In: *AIC Symposium 1988, Colour in Environmental Design*, pp. 10.1–10.5. Winterthur Polytechnic, Winterthur (1988). <https://aic-color.org/resources/Documents/aic1988proc.pdf>. Accessed 24 Mar 2020
15. Sivik, L.: Research on the meanings of color combinations. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. II, pp. 130–133. Grupo Argentino del Color, Buenos Aires (1989). <https://aic-color.org/resources/Documents/aic1989proc2.pdf>
16. Kansaku, J.: The analytic study of affective values of color combinations: a study of color pairs. *Jap. J. Psychol.* **34**, 11–12 (1963)
17. Nayatani, Y.: An analysis of affective values on three-color harmony by the semantic differential method. In: Richter, M. (ed.) *AIC Color 69, Proceedings of the 1st Congress*, pp. 1073–1081. Muster-Schmidt, Göttingen (1970). <https://aic-color.org/resources/Documents/aic1969proc2-reduced.pdf>. <https://aic-color.org/resources/Documents/aic1969proc2-reduced.pdf>
18. Caivano, J.: Color and semiotics: a two-way street. *Color. Res. Appl.* **23**(6), 390–401 (1998)
19. Caivano, J.: Color and cesia as signs: Indexical use of color and cesia in culture and nature. In: *AIC 2014, Interim Meeting of the International Color Association, Proceedings*, pp. 19–28. Amexinc, Asociación Mexicana de Investigadores del Color, México (2015). <https://aic-color.org/resources/Documents/aic2014proc.pdf>. Accessed 24 Mar 2020
20. Luckiesh, M.: *The Language of Color*. Dodd, Mead and Company, New York (1918)
21. Sanz, J.C.: *El lenguaje del color*. Hermann Blume, Madrid (1985)
22. Sanz, J.C.: *Lenguaje del color: sinestesia cromática en poesía y arte visual*. H. Blume, Madrid (2009)
23. Oberascher, L.: The language of colour. In: Nemesics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 137–140. Hungarian National Color Committee, Budapest (1993)

Color Constancy

Bei Xiao

Department of Computer Science, American University, Washington, DC, USA

Definition

Color constancy refers to the ability of the visual system to perceive stable object colors, despite spatial and temporal change, as well as spectral changes, in the illumination. It is a prime example of perceptual constancy [1]. In daily life, humans and animals often count on color constancy to identify, discriminate, and search objects, as well as to judge physical and functional properties of objects. In a related concept, in machine vision, white balancing algorithms in cameras and image acquisition systems are widely used [2, 3]. Color constancy has been an active research topic in the past 100 years. For a thorough understanding of this subject, please see the following recent reviews and papers [4–15].

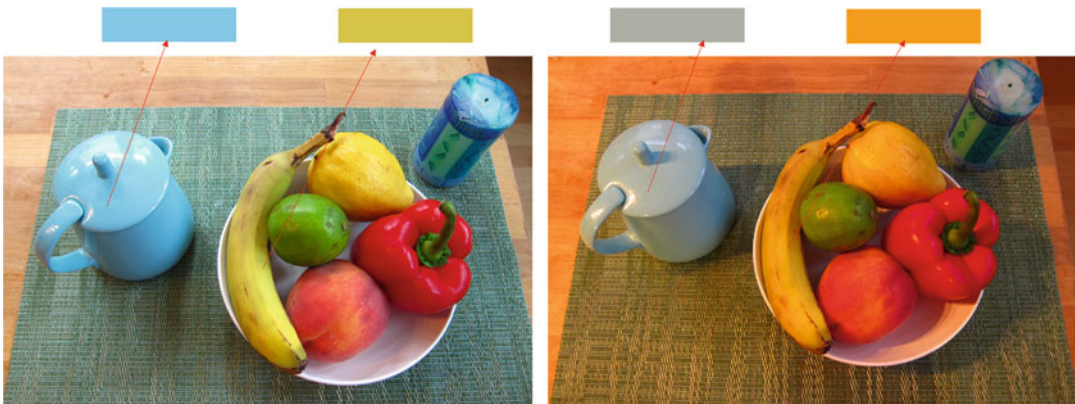
Human Color Constancy

Figure 1 shows photos of the same scene under two illumination conditions. The left photo was taken in the morning when the dominant light source is the daylight, while the right was taken in the evening when the dominant light is tungsten light. If one were in the scene, one would perceive the teapot under both illumination conditions to have the same colors, though individual pixels from the same location from the teapot and the banana in the two photos appear very different (see the patches above the photos). This is also similarly demonstrated by Kitaoka's Twitter account on a red tomato viewed through a green filter. The research question of human color constancy is whether and how the visual system can disentangle the two sources of information, illumination spectrum and surface reflectance, to identify surface colors.

Results of Human Color Constancy Measurements

The visual system generally exhibits high level of color constancy under natural illumination

likely due to its ability to adapt to the environment [16]. However, laboratory measurements of magnitudes of color constancy showed large variation of color constancy measurements in the last 30 years and indicate it is rarely perfect [6, 17–19], for a recent review see [14]. Previous work discovered that instructions and tasks affect the degree of color constancy [20–22]. The articles state that asymmetric matches with appearance-match instructions tend to yield particularly low color constancy indices. The studies that yielded near perfect color constancy (90%) either guaranteed in some way for adaptation [23–26] or used naturalistic scenes where many cues that are available for inferential constancy [27, 28]. Cognitive factors have also been considered in studying color constancy [29–32]. Some objects have characteristic colors, such as bananas are yellow and cucumbers are green. Though previous work shows constancy improved with familiar objects [31], a recent work by Kanematsu [33] suggests the effect of familiar contextual objects on color constancy is small. In Fig. 1, for example, the color of the banana seems stable across the two illuminations. This might be due to our prior knowledge of the object's color.



Color Constancy, Fig. 1 Photographs of the same scene lighted under two illuminants. Left image: Window daylight. Right image: A mixture of indoor artificial lighting sources. The photographs are taken by a Sony Cyber Shot DSC-WX350 compact camera with white-balance turned off. Top panel: the patches show rectangular regions filled with the color from roughly the same locations of the objects in the two images. Bottom panel: the images from which the patches were extracted. Notice

that most objects in the scene remain color constant, though the color patches indicate dramatic change in pixel color. However, material properties such as the specular surface of the teapot do influence the degree of color constancy (most people would agree that the teapot in the left image appears bright blue but grayish blue in the right image). The example also shows familiarity of the object (e.g., banana) helps achieve better constancy



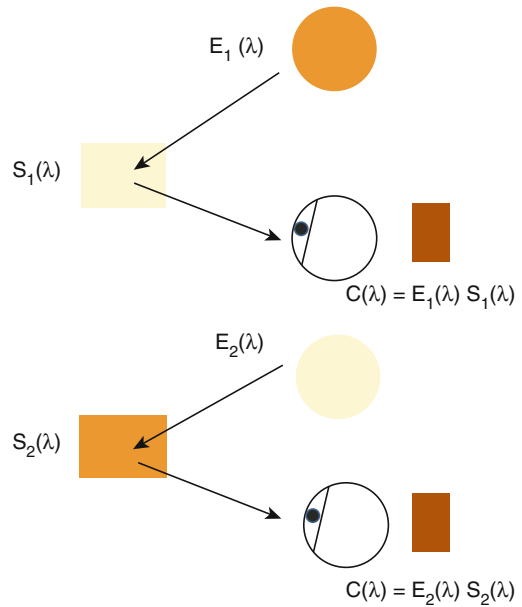
Color Constancy, Fig. 2 Failures of color constancy. Left: A photograph of fruits taken under a monochromatic low-pressure sodium light. Right: The same scene was illuminated by normal broadband light sources. Most observers

won't be able to tell the color of the bell pepper from the left image. But such monochromatic light source is rare in the real world

Under suboptimal lighting (e.g., narrow-band lighting), constancy can also be poor under certain conditions. Figure 2 illustrated a scenario where color constancy fails for most observers. In sum, color constancy is high in natural settings where many cues that allow for a successful inference.

Cues of Color Constancy

Figure 3 demonstrates that color constancy is an ill-posed problem. Reflected light is a function of both illumination and surface reflection. How can the human visual system disambiguate the contribution of the two and achieve constancy? It was shown that humans might use a variety of different image cues about color constancy that are available in different experiments (see reviews [6, 10]. Adaptation and simultaneous local contrast are the major mechanisms that allow for constant color appearances across illuminations. Human observers can also infer color by comparing colors across illuminations that serve as references or anchors. This information can be obtained from depth and motion information [34], specular highlights [35–37] and 3D inter-reflections, shading [38, 39], and memory colors [40–43].



Color Constancy, Fig. 3 Illustration of color constancy. The spectrum of reflected light reaching the eye, $C(\lambda)$ is a wavelength by wavelength product of surface reflectance S and the illuminant E . The problem of color constancy is challenging because different combinations of illumination and surface reflectance can result in the same color signal. This illustration is adapted from Brainard and Radonjić [98]

Constancy Across Surface Colors and Illuminations

Measures of color constancy can change depending on the surface color [18, 31]. Also,

recent works studied the relationship between illumination discrimination thresholds and color constancy and found color constancy depends on the illumination color [44]. They proposed illumination discrimination as a useful new measure of color constancy under different illuminant colors [19]. The consensus of these findings is that observers have a blue bias: they are less sensitive to changes toward the blue direction of the daylight (see [15] for a review and [45–47]), meaning that the measured threshold of illumination discrimination is highest in a bluer chromatic direction. This bias suggests that color constancy might be adapted to daylight illuminations and shadows. Since representation of illumination is closely entangled to representation of surface reflectance at the proximal image, accuracy of estimating illumination of the visual system may directly affect accuracy with which it represents surface reflectance. Accuracy might be higher or lower depending on which direction the illumination change happens. Previous work found that illumination discrimination depends on illumination change direction [48] and the ensemble of surface reflectance in the scene [46].

What Do we Know about Neural Processing of Color Constancy?

Human neuroimaging studies show that responses to color already happen in early areas V1 and V2 [49–51]. Outside area V1 and V2, earlier investigation demonstrated the involvement of area V4 in color perception in monkeys [52, 53] and area IT [54]. However, following these initial works, other studies failed to find clear evidence that V4 was any more involved in color processing than areas V1 and V2 [55]. Recent work in fMRI and micro-electrode recording tried to solve this controversy. The results show that area V4 has a mesoscale organization comprising millimeter-sized color-biased regions, dubbed globs, which have many color-tuned neurons, interspersed with inter-glob regions that have much less color-sensitive neurons (for more details, see [56]). The controversial results are likely caused by recording from neurons at different locations

from area V4. Using physically realistic rendered scenes and multi-voxel pattern analysis, Bannert [57] found only in V1 and V4 surface color representation that is invariant across illumination changes. They also found that along the visual hierarchy, there was a gradient from V1 to V4 to increasingly encode surface color rather than illumination. Even though these studies shine some lights on brain areas involved in color constancy, its neural mechanism is far from understood. One challenge is mapping different sets of psychophysical measurements of color constancy onto different stages of the neural color processing stream. Another challenge is that color constancy might depend on spatial integration of reflected light across surfaces that likely involve many brain areas.

Theories of Human Color Constancy

Can we predict constancy in complex scenes mentioned above? One approach to understand constancy is to explain it using low-level visual mechanisms such as chromatic adaptation. Von Kries proposed that the LMS cone signals are scaled by a multiplicative factor, and at each retinal location, the gains are scaled independently [58, 59]. For each cone class, the gain is set in inverse proportion to the spatial mean of the signals from the cones of the same class. This algorithm is called von Kries adaptation. There are many models developed for color constancy based on cone adaptation models such as the Land's famous retinex theory [60], which has very wide application in camera color balance and can be used to explain human color constancy for the flat-matt-diffuse stimuli. The central principle of the retinex theory is that the lightness values at each pixels are calculated independently for each cone class. For an analysis of the retinex theory and color constancy, see a study by Brainard and Wandell [61]. The adaptation models predict color constancy quite well for flat and mat and diffuse scenes. However, it often fails to predict constancy for rich scenes as mentioned above [16]. We do now know how to obtain the values of the multiplicative gain from images,

which contain spatially rich information. Another approach is to develop computational models that could be used to predict human color constancy in complex scenes. The *equivalent illuminant model* assumed that the only error the visual system makes in an attempt to achieve constancy is that it misestimates the scene illuminant [62]. In a Bayesian framework, Brainard and Freeman [63] constructed prior distribution that describes the probability of illuminants and surfaces in the world, and then estimates the illuminant from the posterior distribution conditioned on the image intensity data. Further studies applied the similar Bayesian model to predict the degree of human color constancy across different manipulations and connect the variation in constancy to the prior distribution of the illuminant [64, 65].

#TheDress and Individual Variations in Constancy

In 2015, on the internet a photo of dress (#theDress) became a viral social media sensation with viewers disagreeing whether the dress is blue or black, or white and gold. A large proportion of observers saw the colors of the body of #theDress as blue and the lace black, while some others saw them as white and gold [66–68]. Even though there is currently no consensus on why the dress elicits such a discordant color perception among the viewers, there have been many scientific articles published proposing the role of illumination estimation on the individual perception [69–72]. It was found that the perceived colors of both body of the dress and the lace as well as the physical pixels of the body of the dress and the lace were distributed in a continuous range of hues close to the daylight locus [69]. This alignment of the pixel color signals in the image makes it possible to attribute the color variation either to the fabric or to the lighting conditions. The assumption about the illumination is different between observers. Some observers implicitly assumed the dress to be in the shadow, implying bluish-dark lighting, and they saw #theDress as white and gold. In contrast, some observers believed the lighting to be white and bright and saw #theDress as blue and black. The variation of the pixel colors

along the daylight locus could contribute to the magnitude of individual perceived differences. In addition to explain the phenomena in the light of illumination estimation, there is a strong interaction between illumination, material properties, pigment of the fabric material, and color constancy. Fabrics interact with light in complex ways, and this might affect the degree of color constancy [73]. Overall, the phenomena of #TheDress intensified the discussion of the plausibility that the neural mechanisms underlying color constancy are optimized for illuminations along the daylight locus [74]. It also raises attention to the role of the interindividual variability in color constancy.

Machine Color Constancy

While the human visual system is equipped with good color constancy in most situations, the digital camera has to rely on color-balancing algorithm to discount the illumination effect and extract the invariable object color. This process is also called color balance or white balance. “Color constancy” means recovering the vertical world behind an image, while “white balance” means giving the image pleasant appearance by discounting the effect of illumination that is consistent with some aesthetic norm. The vast majority of computational color constancy research is focused on illumination estimation. The methods can be divided into two categories: statistics-based method and learning-based.

The best-known statistical method is the gray-world theory, which assumes that the average reflectance of a scene is gray [75]. Some other similar algorithms include white patch and max-RGB [76], and shade of gray [77]. The gray-world algorithm will fail if the average reflectance is not achromatic or if there is a large uniform colored surface in the scene. The incorporation of higher-order statistics in terms of image-derivatives is proposed, where a framework called gray edge is presented [78].

Learning-based methods estimate the color of illuminant using a method learned from training data with labeled ground-truth. Chakrabarti [79] et al. go beyond statistics of per-pixel colors and

model the spatial dependencies between pixels by decomposing the input images into spatial sub-bands, and then model the color statistics separately in each sub-band. Forsyth [80] introduced the gamut mapping method. It is based on the assumption that only a limited set of colors (canonical gamut) can occur under a given illuminant. The model learns a model based on training images (the canonical gamut) and estimate the illuminant based on the input features. Inspired by Brainard [63], Rosenberg et al. [81] introduced a Bayesian model of color constancy utilizing a non-Gaussian probabilistic model of the image formation process and demonstrated that it can outperform the gamut-mapping algorithm. Gehler et al. [82] extended the Bayesian algorithm using a new dataset, which allows the algorithm to learn more precise priors of the illuminations. Barron [83] developed the Fast Fourier color constancy algorithm that can be readily used on a mobile device with 700 frames per second. The algorithm is based on estimating a complete posterior distribution of illumination instead of a single illuminant.

Many of these models were trained on hand-crafted features extracted from the input image [84–87]. For example, Bianco et al. [88] proposed a novel model to estimate illuminant from color statistics extracted from faces automatically detected in the image. It takes advantage of the fact that the skin color tends to cluster in the color space, which provides a valid cue to estimate the illuminant. Inspired by human visual system mechanisms, a recent study by Gao et al. [89] built a physiologically based color constancy model that imitated the double-opponent mechanism of the visual system. The illuminant is estimated by searching for the maxima of the separate RGB channels of the responses of double-opponent cells in the RGB space.

The most recent learning-based models learn features by using deep convolutional neural networks (CNNs). Bianco et al. [90] was the first to use a CNN to learn discriminant features for the illuminant estimation task and similar deep learning-based color constancy algorithms were developed [91–94]. The challenging of the deep learning-based methods is the requirement of large-scale datasets with ground-truth illuminants

and the current largest datasets for color constancy are much smaller than those of object recognition. To solve this issue, several unsupervised learning methods in the neural network framework has been proposed that does not need such as labeled database. Qian et al. [95] proposed a statistical color constancy method that relies on a novel gray pixel detection followed by mean shift clustering. Bianco et al. [96] trained networks to detect achromatic pixels after the images are converted to grayscale as a proxy of illuminant estimation and then used this estimation to perform the white balancing. A new work avoided to explicitly estimate illumination but proposed an algorithm that corrects a camera image that has been improperly white-balanced [97]. The data-driven approach takes an incorrectly white-balanced sRGB image and computes a mapping that can transform the input image's colors to appear as if the white balance was correctly applied.

Future Directions

Color constancy is an important and practical problem in both human and computer vision. Color constancy provides an excellent model system to understand how the visual system solves ambiguity. Significant progress has been made to understand color constancy of simple scenes. A big challenge is how to extend the theoretical model of color constancy for simple scenes to predict human performance in rich scenes and explain individual variabilities. How color is encoded in the cortex is still a major scientific question. Mapping different types of psychophysical measurements of color constancy onto different neural processing stages and streams is challenging. The current experimental evidence suggests that color information is encoded throughout the visual cortex and it is possible that color constancy is variously represented.

Major breakthroughs in computational color constancy algorithms have been made thanks to development of large datasets and machine learning methods. However, it is unknown whether or not the human visual system uses similar algorithms to achieve constancy and how to improve white balance algorithm based on human color

constancy. How to bridge the two fields is a promising endeavor for future research.

Cross-References

► Environmental Influences on Color Vision

References

- Walsh, V., Kulikowski, J.: *Perceptual Constancy: why Things Look as they Do*. Cambridge University Press (1998)
- Helmholtz, H.V.: LXIII. On integrals of the hydrodynamical equations, which express vortex-motion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. **33**(226), 485–512 (1867)
- Shepard, R.N.: Perceptual-cognitive universals as reflections of the world. *Behav. Brain Sci.* **24**(4), 581–601 (2001)
- Brainard, D.H., Maloney, L.T.: Perception of color and material properties in complex scenes. *J. Vis.* **4**(9), i–i (2004)
- Zaidi, Q., Clifford, C., Rhodes, G.: *The Role of Adaptation in Colour Constancy*. Oxford University, Oxford (2005)
- Smithson, H.E.: Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **360**(1458), 1329–1346 (2005)
- Hurlbert, A.: Colour constancy. *Curr. Biol.* **17**(21), R906–R907 (2007)
- Shevell, S.K., Kingdom, F.A.: Color in complex scenes. *Annu. Rev. Psychol.* **59**, 143–166 (2008)
- Maloney, L.T., Brainard, D.H.: Color and material perception: achievements and challenges. *J. Vis.* **10**(9), 19–19 (2010)
- Foster, D.H.: Color constancy. *Vis. Res.* **51**(7), 674–700 (2011)
- Stockman, A., Brainard, D.H.: Color vision mechanisms. *OSA Handbook of Optics*. 11–11 (2010)
- Bianco, S., Schettini, R.: Computational color constancy. In: *3rd European Workshop on Visual Information Processing*, pp. 1–7. IEEE (2011)
- Olkkonen, M., Ekroll, V.: Color constancy and contextual effects on color appearance. In: *Human Color Vision*, pp. 159–188. Springer, Cham (2016)
- Witzel, C., Gegenfurtner, K.R.: Color perception: objects, constancy, and categories. *Ann. Rev. Vis. Sci.* 475–499 (2018)
- Hurlbert, A.: Challenges to color constancy in a contemporary light. *Curr. Opin. Behav. Sci.* **30**, 186–193 (2019)
- Kraft, J.M., Brainard, D.H.: Mechanisms of color constancy under nearly natural viewing. *Proc. Natl. Acad. Sci.* **96**(1), 307–312 (1999)
- Maloney LT. Physics-based approaches to modeling surface color perception. *Color Vision: From Genes to Perception*. 1999; p. 387–416
- Witzel, C., van Alphen, C., Godau, C., O'Regan, J.K.: Uncertainty of sensory signal explains variation of color constancy. *J. Vis.* **16**(15), 8–8 (2016)
- Aston, S., Radonjić, A., Brainard, D.H., Hurlbert, A.C.: Illumination discrimination for chromatically biased illuminations: implications for color constancy. *J. Vis.* **19**(3), 15–15 (2019)
- Bäuml, K.H.: Simultaneous color constancy: how surface color perception varies with the illuminant. *Vis. Res.* **39**(8), 1531–1550 (1999)
- Radonjić, A., Cottaris, N.P., Brainard, D.H.: Color constancy supports cross-illumination color selection. *J. Vis.* **15**(6), 13–13 (2015)
- Radonjić, A., Brainard, D.H.: The nature of instructional effects in color constancy. *J. Exp. Psychol. Hum. Percept. Perform.* **42**(6), 847 (2016)
- Brainard, D.H.: Color constancy in the nearly natural image. 2. Achromatic loci. *JOSA A*. **15**(2), 307–325 (1998)
- Hansen, T., Walter, S., Gegenfurtner, K.R.: Effects of spatial and temporal context on color categories and color constancy. *J. Vis.* **7**(4), 2–2 (2007)
- Murray, I.J., Daugirdiene, A., Stanikunas, R., Vaitkevicius, H., Kulikowski, J.: Cone contrasts do not predict color constancy. *Vis. Neurosci.* **23**(3–4), 543–547 (2006)
- Rinner, O., Gegenfurtner, K.R.: Cone contributions to colour constancy. *Perception*. **31**(6), 733–746 (2002)
- Olkkonen, M., Allred, S.R.: Short-term memory affects color perception in context. *PLoS One*. **9**(1), (2014)
- Xiao, B., Hurst, B., MacIntyre, L., Brainard, D.H.: The color constancy of three-dimensional objects. *J. Vis.* **12**(4), 6–6 (2012)
- Hurlbert, A.C., Ling, Y.: If it's a banana, it must be yellow: the role of memory colors in color constancy. *J. Vis.* **5**(8), 787–787 (2005)
- Olkkonen, M., Hansen, T., Gegenfurtner, K.R.: Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J. Vis.* **8**(5), 13–13 (2008)
- Olkkonen, M., Witzel, C., Hansen, T., Gegenfurtner, K.R.: Categorical color constancy for real surfaces. *J. Vis.* **10**(9), 16–16 (2010)
- Smet, K., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**(3), 192–200 (2011)
- Kanematsu, E., Brainard, D.H.: No measured effect of a familiar contextual object on color constancy. *Color Res. Appl.* **39**(4), 347–359 (2014)
- Werner, A.: Color constancy improves, when an object moves: high-level motion influences color perception. *J. Vis.* **7**(14), 19–19 (2007)
- Yang, J.N., Maloney, L.T.: Illuminant cues in surface color perception: tests of three candidate cues. *Vis. Res.* **41**(20), 2581–2600 (2001)

36. Granzier, J.J., Vergne, R., Gegenfurtner, K.R.: The effects of surface gloss and roughness on color constancy for real 3-D objects. *J. Vis.* **14**(2), 16–16 (2014)
37. Lee, R.J., Smithson, H.E.: Low levels of specularly support operational color constancy, particularly when surface and illumination geometry can be inferred. *JOSA A.* **33**(3), A306–A318 (2016)
38. Bloj, M.G., Kersten, D., Hurlbert, A.C.: Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature.* **402**(6764), 877–879 (1999)
39. Hedrich, M., Bloj, M., Ruppertsberg, A.I.: Color constancy improves for real 3D objects. *J. Vis.* **9**(4), 16–16 (2009)
40. Jin, E.W., Shevell, S.K.: Color memory and color constancy. *JOSA A.* **13**(10), 1981–1991 (1996)
41. Ling, Y., Hurlbert, A.: Role of color memory in successive color constancy. *JOSA A.* **25**(6), 1215–1226 (2008)
42. Witzel, C., Hansen, T.: Memory effects on color perception. In: Elliot, A.J., Franklin, A., Fairchild, M.D. (eds.) *Handbook of Color Psychology*, pp. 639–640. Cambridge University Press, Cambridge, UK (2015)
43. Granzier, J.J., Gegenfurtner, K.R.: Effects of memory colour on colour constancy for unknown coloured objects. *i-Perception.* **3**(3), 190–215 (2012)
44. Álvaro, L., Linhares, J.M., Moreira, H., Lillo, J., Nascimento, S.M.: Robust colour constancy in red-green dichromats. *PLoS One.* **12**(6), e0180310 (2017)
45. Winkler, A.D., Spillmann, L., Werner, J.S., Webster, M.A.: Asymmetries in blue–yellow color perception and in the color of ‘the dress’. *Curr. Biol.* **25**(13), R547–R548 (2015)
46. Radonjic, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N.P., et al.: Illumination discrimination in real and simulated scenes. *J. Vis.* **16**(11), 2–2 (2016)
47. Weiss, D., Witzel, C., Gegenfurtner, K.: Determinants of colour constancy and the blue bias. *i-Perception.* **8**(6), 2041669517739635 (2017)
48. Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G.D., Hurlbert, A.: Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS One.* **9**(2), (2014)
49. Bartels, A., Zeki, S.: The architecture of the colour Centre in the human visual brain: new results and a review. *Eur. J. Neurosci.* **12**(1), 172–193 (2000)
50. Beauchamp, M.S., Haxby, J.V., Jennings, J.E., DeYoe, E.A.: An fMRI version of the Farnsworth–Munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex.* **9**(3), 257–263 (1999)
51. Engel, S.A., Glover, G.H., Wandell, B.A.: Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex.* **7**(2), 181–192 (1997)
52. Zeki, S.M.: Colour coding in rhesus monkey prestriate cortex. *Brain Res.* **53**(2), 422–427 (1973)
53. Zeki, S.: Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wave-lengths and colours. *Neuroscience.* **9**(4), 741–765 (1983)
54. Komatsu, H.: Mechanisms of central color vision. *Curr. Opin. Neurobiol.* **8**(4), 503–508 (1998)
55. Schein, S.J., Desimone, R.: Spectral properties of V4 neurons in the macaque. *J. Neurosci.* **10**(10), 3369–3389 (1990)
56. Conway, B.R., Eskew Jr., R.T., Martin, P.R., Stockman, A.: A tour of contemporary color vision research. *Vis. Res.* **151**, 2–6 (2018)
57. Bannert, M.M., Bartels, A.: Invariance of surface color representations across illuminant changes in the human cortex. *NeuroImage.* **158**, 356–370 (2017)
58. von Kries J. Chromatic Adaptation, *Festschrift der Albrecht-Ludwig-Universität*; 1902
59. Webster, M.A., Mollon, J.: Adaptation and the color statistics of natural images. *Vis. Res.* **37**(23), 3283–3298 (1997)
60. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**(6), 108–129 (1977)
61. Brainard, D.H., Wandell, B.A.: Analysis of the retinex theory of color vision. *JOSA A.* **3**(10), 1651–1661 (1986)
62. Brainard, D.H., Brunt, W.A., Speigle, J.M.: Color constancy in the nearly natural image. 1. Asymmetric matches. *JOSA A.* **14**(9), 2091–2110 (1997)
63. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. *JOSA A.* **14**(7), 1393–1411 (1997)
64. Brainard, D.H., Longère, P., Delahunt, P.B., Freeman, W.T., Kraft, J.M., Xiao, B.: Bayesian model of human color constancy. *J. Vis.* **6**(11), 10–10 (2006)
65. Brainard D, Gazzaniga M. Bayesian approaches to color vision. *The cognitive neurosciences.* 2009.;4 MIT Press, Cambridge, MA:395–408
66. Lafer-Sousa, R., Hermann, K.L., Conway, B.R.: Striking individual differences in color perception uncovered by ‘the dress’ photograph. *Curr. Biol.* **25**(13), R545–R546 (2015)
67. Brainard, D.H., Hurlbert, A.C.: Colour vision: understanding# TheDress. *Curr. Biol.* **25**(13), R551–R554 (2015)
68. Dixon, E.L., Shapiro, A.G.: Spatial filtering, color constancy, and the color-changing dress. *J. Vis.* **17**(3), 7–7 (2017)
69. Gegenfurtner, K.R., Bloj, M., Toscani, M.: The many colours of ‘the dress’. *Curr. Biol.* **25**(13), R543–R544 (2015)
70. Wallisch, P.: Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: “the dress”. *J. Vis.* **17**(4), 5–5 (2017)
71. Uchikawa, K., Morimoto, T., Matsumoto, T.: Understanding individual differences in color appearance of “# TheDress” based on the optimal color hypothesis. *J. Vis.* **17**(8), 10–10 (2017)
72. Lafer-Sousa, R., Conway, B.R.: # TheDress: categorical perception of an ambiguous color image. *J. Vis.* **17**(12), 25–25 (2017)

73. Zhao, S., Jakob, W., Marschner, S., Bala, K.: Building volumetric appearance models of fabric using micro CT imaging. *ACM Trans. Graph. (TOG)*. **30**(4), 1–10 (2011)
74. Lafer-Sousa, R., Liu, Y.O., Lafer-Sousa, L., Wiest, M. C., Conway, B.R.: Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. *JOSA A*. **29**(5), 657–670 (2012)
75. Buchsbaum, G.: A spatial processor model for object colour perception. *J. Frankl. Inst.* **310**(1), 1–26 (1980)
76. Funt, B., Shi, L.: The rehabilitation of maxrgb. In: *Color and Imaging Conference*, vol. 1, pp. 256–259. Society for Imaging Science and Technology (2010)
77. Finlayson, G.D., Hordley, S.D., Hubel, P.M.: Color by correlation: a simple, unifying framework for color constancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1209–1221 (2001)
78. Van De Weijer, J., Gevers, T., Gijzenij, A.: Edge-based color constancy. *IEEE Trans. Image Process.* **16**(9), 2207–2214 (2007)
79. Chakrabarti, A., Hirakawa, K., Zickler, T.: Color constancy with spatio-spectral statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(8), 1509–1519 (2011)
80. Forsyth, D.A.: A novel algorithm for color constancy. *Int. J. Comput. Vis.* **5**(1), 5–35 (1990)
81. Rosenberg, C., Ladsariya, A., Minka, T.: Bayesian color constancy with non-gaussian models. In: *Advances in Neural Information Processing Systems*, pp. 1595–1602 (2004)
82. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE (2008)
83. Barron, J.T., Tsai, Y.T.: Fast fourier color constancy. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–894 (2017)
84. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Classification-based color constancy. *International Conference on Advances in Visual Information Systems*. Springer, 104–113 (2008)
85. Gijzenij, A., Gevers, T.: Color constancy by local averaging. In: *14th International Conference of Image Analysis and Processing-Workshops (ICIAPW 2007)*, pp. 171–174. IEEE (2007)
86. Van De Weijer, J., Schmid, C., Verbeek, J.: Using high-level visual information for color constancy. In: *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8. IEEE (2007)
87. Gijzenij, A., Gevers, T.: Color constancy using natural image statistics and scene semantics. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(4), 687–698 (2010)
88. Bianco, S., Schettini, R.: Color constancy using faces. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 65–72. IEEE (2012)
89. Gao, S., Yang, K., Li, C., Li, Y.: A color constancy model with double-opponency mechanisms. *Proceedings of the IEEE international conference on computer vision*. 929–936 (2013)
90. Bianco, S., Cusano, C., Schettini, R.: Color constancy using CNNs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 81–89 (2015)
91. Lou, Z., Gevers, T., Hu, N., Lucassen, M.P.: Color Constancy by deep learning. In: Xianghua Xie, M.W. J., Tam, G.K.L. (eds.) *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 76.1–76.12. BMVA Press (2015). <https://doi.org/10.5244/C.29.76>
92. Shi, W., Loy, C.C., Tang, X.: Deep specialized network for illuminant estimation. *European Conference on Computer Vision*. Springer, 371–387 (2016)
93. Hu, Y., Wang, B., Lin, S.: Fc4: Fully convolutional color constancy with confidence-weighted pooling. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4085–4094 (2017)
94. Bianco, S., Cusano, C., Schettini, R.: Single and multiple illuminant estimation using convolutional neural networks. *IEEE Trans. Image Process.* **26**(9), 4347–4362 (2017)
95. Qian, Y., Chen, K., Nikkanen, J., Kamarainen, J.K., Matas, J.: Recurrent color constancy. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5458–5466 (2017)
96. Bianco, S., Cusano, C.: Quasi-unsupervised color constancy. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12212–12221 (2019)
97. Afifi, M., Price, B., Cohen, S., Brown, M.S.: When color constancy goes wrong: correcting improperly white-balanced images. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 1535–1544 (2019)
98. Brainard, D., Radonjić, A.: Color constancy. In: *The New Visual Neurosciences*, pp. 545–556. MIT Press (2014)

Color Constancy and Categories

Maria Olkkonen

Department of Psychology and Logopedics,
Faculty of Medicine, University of Helsinki,
Helsinki, Finland

Department of Psychology, Durham University,
Durham, UK

Synonyms

Categorical color constancy; Color categorization; Color classification; Color identification; Color naming; Color sorting

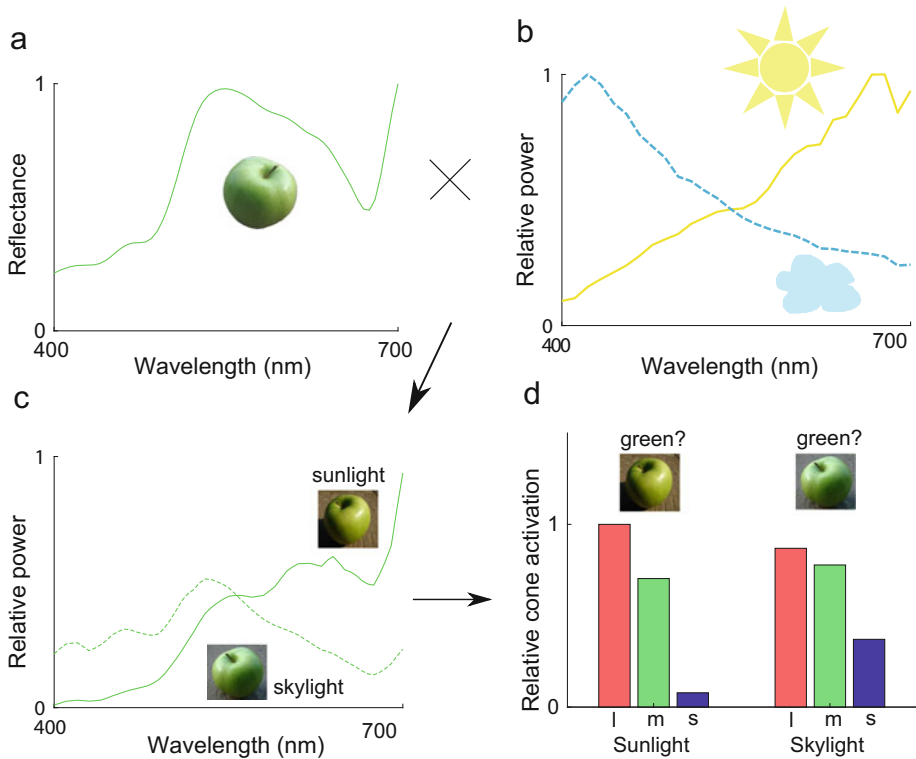
Definition

Color constancy is defined here as the ability to identify surface color across variations in ambient illumination or scene context. Presuming that an important function of color vision is to help identify objects and object properties based on surface color (e.g., ripe fruit, fresh fish), stable appearance is not a necessary requirement, as long as the percept is sufficiently constant for color identification. One aspect of color perception that may be relevant here is the ability to classify colors into a small number of categories; it may be sufficient for color identification across illumination changes to maintain constancy of color categories.

Color Constancy

The Problem

Color perception is a computational challenge because the light reaching the eye confounds information about the stable surface property that we care about – surface reflectance – and the ambient illumination. This is compounded by the fact that the light spectrum is encoded into three quantities in a trichromatic retina, that is, in the activations of the long-wavelength-sensitive, middle-wavelength-sensitive, and short-wavelength-sensitive cone photoreceptors (Fig. 1; [1]). It is not possible, without additional constraints, to recover the spectral reflectance from the light signal because of an infinite number



Color Constancy and Categories, Fig. 1 The problem of color constancy. (a) Each surface has a spectral reflectance, which indicates how much light is reflected at each wavelength of the visible light spectrum. Reflectance is a stable object property, which correlates with perceived surface color. (b) Each illumination has a power spectrum which indicates how much energy it has at each wavelength. The panel shows the power spectra of two daylight

illuminants, a bluish skylight and a yellowish sunlight. (c) The power spectrum of the light that reflects to the eye from a surface is a product of its reflectance with the illuminant power spectrum. (d) The light signal is coded with three quantities in a trichromatic retina – the l-, m-, and s-cone photoreceptors. Perceived color categories may remain more stable under illumination changes than color appearance

of possible combinations of spectral reflectance and illumination. The severity of this problem may surprise the reader considering that we rarely notice any significant “errors” in color perception in our everyday lives, except perhaps under extreme or unusual illumination changes. This is despite the fact that the reflected light from a surface varies considerably even under moderate illumination changes (see Fig. 1c–d).

Measuring Color Constancy

Color constancy has traditionally been characterized by measuring the appearance of small stimuli on uniform backgrounds of varying chromaticity (e.g., [4]). The rationale is as follows: if the observer assumes that the background chromaticity equals the illumination chromaticity, by measuring the color appearance of the central stimulus on different backgrounds, we can calculate the extent to which the observer is compensating for the implied illumination. This can be achieved with two widely used tasks: achromatic settings and asymmetric matching. In the first task, the observer is asked to adjust a stimulus embedded in a background to appear achromatic, that is, neutral gray, while the surround is changed between sessions. Color constancy is then quantified as the change in the observers’ setting in relation to the change in the background (as a proxy of the illumination). In asymmetric matching, the observer is asked to match two stimuli embedded in two adjacent (simultaneous matching) or successive (successive matching) backgrounds so that they appear identical. Color constancy is quantified as the relationship between the test and match chromaticity in relation to the difference in illumination.

The achromatic setting task is usually intuitive for observers, but its scope is limited as it only measures the constancy of the achromatic point. Asymmetric matches have the benefit over achromatic settings that they can be collected for many different color pairs in color space, but observers often report that matching stimuli across two different contexts is difficult. Furthermore, measuring many points in color space is time consuming, limiting the number of stimuli used.

Critically, asymmetric matches depend on instructions (for a review, see [2]): observers show higher constancy when they are instructed to match the surfaces as if they were cut from the same piece of paper than if they are just instructed to make an appearance match. This may be explained by the dual nature of color experience: we can appreciate both the stability and the variability in surface colors when the illumination changes, and the two instructions might be tapping into these two different modes of color experience.

Finally, both tasks are unnatural: the goal of color constancy in real life is not to make two stimuli to appear identical or gray but to identify surface colors robustly. These shortcomings of achromatic adjustment and asymmetric matching have received more attention recently, and methods that are less dependent on instructions and observer strategies are increasingly used.

Zaidi and colleagues [8] developed a paradigm that resembles real-life color tasks. In these experiments, observers are shown two objects in one illumination context and two other objects in an adjacent illumination context. Three of the objects have the same surface color, while one is different. The observers are asked to identify the odd-one-out object. One salient advantage of this is that color need not be mentioned, making this task possibly less sensitive to observer bias or demand characteristics. To pick the odd object out, the observer has to first decide which context contains the two stimuli that differ in surface color and then decide which of those two stimuli is different to the two in the other context. This task is potentially more intuitive for observers than the traditional methods because observers are not required to make matches across contexts, and furthermore, the paradigm allows measuring contextual effects on color appearance and color discrimination within the same trial. This paradigm has revealed that observers are good but not perfect at picking the correct object, and the pattern of errors gives important insight into the strategies used by observers.

Recently, Radonjić and colleagues developed a color selection task with a similar principle to that of Zaidi [2]. Their use of rendered stimuli affords

more flexibility in the choice of stimuli and thus more accurate measurements of color constancy. Analyzing the selection data with a variant of maximum likelihood difference scaling, Radonjić et al. showed that observers were largely color constant with complex, 3D stimuli, but quite poor with simple, 2D stimuli. They further found that color constancy in a naturalistic block-sorting task was robust to changes in scene complexity (i.e., number of surfaces in the scene) and local contrast. Finally, they showed that although task instructions – whether observers were asked to match appearance or surface reflectance – affected color constancy in both simple and complex scenes, the effects were much smaller in complex scenes, and color constancy was overall much higher. This indicates that results from simple scenes and unnatural tasks cannot necessarily be generalized to natural scenes.

Constancy and Categories

Color Categorization as a Method to Study Color Constancy

Although we can discriminate millions of colors [6], we also naturally use a small number of color names to describe object color [14]. Color categories are useful when communicating about colors with others, especially considering the physiological differences across individuals in the front end of vision [9] likely to cause differences in how specific colors are perceived. In addition to being beneficial for social communication, color categorization might support color constancy. Imagine trying to pick a ripe tomato in the supermarket. The tomatoes might all have different shades varying between orange and red, and as the illumination is unknown, you have to guess which tomatoes are truly ripe. It is probably safest to pick the tomatoes that you would call red rather than orange. This makes color categorization a useful tool for characterizing real-life color constancy.

In categorical color constancy experiments, observers are asked to name the color of each of a large number of real or rendered stimuli under different real or simulated illuminations. The

stimuli are often sampled from a collection of carefully produced surfaces, such as the chips in the Munsell collection. The reflectance characteristics of these surfaces are known, and as long as the illumination spectrum is also known, the spectral power distribution of the light reflecting off the surfaces to the observer's eyes can be easily calculated as the product of the surface spectral reflectance and the illumination power spectrum. Such chips can also be rendered on a calibrated monitor display, if more control in the display of the stimuli and the illumination are required. Categorical color constancy studies often employ one neutral illuminant as a baseline, such as the standard CIE D65 illuminant which is a neutral daylight with a color temperature of 6500 K. In addition, there are often 2–4 chromatic illuminants sampled symmetrically around the neutral illuminant either toward yellowish and bluish or toward greenish and reddish. We can then ask how stable are the color names given to each chip under the different illuminants. This can be quantified by, for instance, calculating a color naming index, which is the number of “same” classifications across all illuminants divided by the total number of illuminants. 0 indicates no categorical color constancy – all color names are different – and 1 indicates perfect categorical color constancy – all names are the same. This is a useful summary measure, but it averages over all illumination changes from neutral. If we want to compare individual illumination changes, we can calculate a similar index separately for each illumination change from neutral. Averaged over all color samples, the index provides a measure of overall color constancy for each illumination.

In addition to analyzing naming consistency, category stability can further be assessed by fitting category boundaries to the naming data and analyzing the transformations in the boundaries caused by illumination changes. This is a more continuous measure of category stability that allows for analyzing how the *structure* of color naming space changes with the illumination. Finally, the centroids of each color category can be defined and their shifts analyzed. This is a measure of color constancy that can readily be

compared with the more traditional achromatic adjustment and asymmetric matching measures.

Measuring Chromatic Adaptation with Color Naming

Early investigations of color category stability characterized the effect of chromatic adaptation on color naming. Chromatic adaptation is a low-level mechanism of color constancy, whereby the visual system adapts to the overall chromaticity in the visual field, causing measurable changes in color appearance [10]. These changes can be quantified and taken as a measure of color constancy.

An early study asked observers to name a set of color stimuli spanning the visible spectrum with a small number of color names (red, green, blue, or yellow, or their combinations) under different adapting chromaticities [5]. Chromatic adaptation affected the frequency distribution of color names: under adaptation to a reddish light, the frequency of “green” increased and “red” decreased; under a bluish light, the frequency of “yellow” increased and “blue” decreased; and so forth. Functionally, this pattern is consistent with the “neutral point” of the visual system shifting toward the adapting chromaticity, making previously chromatic stimuli look achromatic (gray) and previously achromatic stimuli look tinted in the color opposite from the adapting color. The legacy of this early work lies in demonstrating that the effect of chromatic adaptation (and by extension, color constancy) can be measured with a method that is more intuitive for observers and allows for measuring the stability of more than one region of color space (also see [3]).

Although the early chromatic adaptation studies were influential, the relationship to color constancy in real life is tenuous because of how the stimuli were defined: as points in wavelength or cone contrast space and not as surfaces. That chromatic adaptation has an effect on how these kinds of stimuli are categorized indicates that the visual system is adapting to the illumination in a way that would produce partial color constancy, but there is no obvious way to determine an upper bound for adaptation, or “perfect” constancy. If the stimuli were defined in terms of identities, for

example, by their surface reflectance, constancy could be characterized as the consistency in classifying a surface in the same color category under each of the test illuminations, as explained above. In addition to measuring the effect of illumination on color naming, one could also calculate within-observer and across-observer color naming reliability for the baseline illumination.

Color Naming with Real or Simulated Surfaces

More recently, studies into categorical color constancy have used either real or simulated color samples. Troost and de Weert [13] used a large set of stimuli varying in hue, saturation, and lightness simulated under a neutral and 4 chromatic illuminations and asked observers to categorize each stimulus into 1 of 12 predetermined categories (roughly corresponding to the basic color terms). They quantified categorical color constancy in two principal ways: first, they calculated a consistency index across the five illuminations; second, they calculated the constancy of the centroids of each color category under each illumination change from neutral. They found, on average, high naming consistency and relatively high constancy from the centroids. They also ran an asymmetric matching experiment with two sets of instructions: observers were asked to either make a paper match or an appearance match. Categorical color constancy was more comparable to paper matches than to appearance matches, which is consistent with the notion that color categorization pertains to object properties rather than to “proximal” stimulus properties. Speigle and Brainard [12] also compared asymmetric matching and color naming as measures of color constancy with a systematic analysis whereby they predicted the color naming data from the matching data. They found that the agreement between matching and naming data for a given observer was comparable to the agreement between two observers for color naming. Smithson and Zaidi [11] used opponent color classification (similar to that used in [3]) to study different hypotheses about color constancy mechanisms: adaptation to the local mean, adaptation to the global mean, and temporal integration. They found that color constancy was nearly perfect

under illumination changes from a bluish to a yellowish daylight; that constancy was not determined by the global mean over the scene; and that observers were able to extract information about the illumination through spatially local temporal integration.

One distinct benefit of quantifying color constancy with color categorization is that it allows researchers to assess the stability of different regions of color space under illumination changes and to compare this to category stability across different observers to gain insight into the communicative aspects of color constancy. Olkkonen, Witzel, Hansen, and Gegenfurtner [7] had observers sort Munsell chips into different basic color categories under one neutral and four chromatic illuminations. They evaluated category sorting consistency across repetitions for the same illumination (within-observer), across different illuminations for the same observer (across-illumination), and across different observers for the same illumination (across-observer). They showed that categorical color constancy was virtually as high as within-observer consistency; in other words, an illumination change did not cause any more variability in color naming than just doing the sorting task twice under a neutral illumination. Moreover, the same color samples remained stable under an illumination change for a given observer and across-observers for the neutral illumination. This provides evidence for the notion that color naming consistency across speakers of a given language is linked to the perceptual stability of color categories under illumination changes.

Summary and Outlook

In summary, categorical color constancy has been found to be better than so-called appearance constancy, in line with the intuition that color names remain more stable than color appearance under illumination changes. Although much has been learned about both appearance constancy and category constancy in the laboratory, important open questions remain, to which color categorization might provide further insight. Do results obtained

in the laboratory hold in naturalistic viewing situations? Does color constancy depend on other object properties such as shape and material? Which visual cues are used in complex scenes to estimate surface color? An important endeavor is to formulate a computational framework in which existing results can be integrated and from which new predictions may be derived and tested. Bayesian models show much promise [1], but other frameworks are plausible. Fortunately, the methods employed to characterize color constancy are moving toward more naturalistic paradigms, which are more likely than traditional approaches to reveal real-world color constancy mechanisms. For now, researchers are limited to using either calibrated displays or real scenes, both of which have their own issues. One promising avenue is using virtual reality to combine the advantages of controllable displays and realistic scenes, once the characterization and calibration of VR devices for color display have matured.

Cross-References

- [Color Constancy](#)
- [Color Order Systems](#)
- [Interactions of Color and Light](#)
- [Metamerism](#)
- [Psychological Color Space and Color Terms](#)
- [Simultaneous Color Contrast](#)
- [Spectral Power Distribution](#)

References

1. Allred, S.R.: Visual Experience: Sensation, Cognition, and Constancy (ed: Hatfield, G., Allred, S.R.). Oxford University Press, Oxford (2012)
2. Brainard, D.H., Cottaris, N.P., Radonjić, A.: The perception of color and material in naturalistic tasks. *Interface Focus*. **8**, 20180012 (2018)
3. Chichilnisky, E.J., Wandell, B.A.: Trichromatic opponent color classification. *Vis. Res.* **39**, 3444–3458 (1999)
4. Helson, H.: Fundamental problems in color vision. I. The principle governing changes in hue, saturation and lightness of non-selective samples in chromatic illumination. *J. Exp. Psychol.* **23**, 439–476 (1938)

5. Jacobs, G.H., Gaylord, H.A.: Effects of chromatic adaptation on color naming. *Vis. Res.* **7**, 645–653 (1967)
6. Linhares, J.M., Pinto, P.D., Nascimento, S.M.: The number of discernible colors in natural scenes. *J. Opt. Soc. Am. A* **25**, 2918–2924 (2008)
7. Olkkonen, M., Witzel, C., Hansen, T., Gegenfurtner, K.R.: Categorical color constancy for real surfaces. *J. Vis.* **10**, 9.1–9.22. ISSN: 15347362 (2010)
8. Robilotto, R., Zaidi, Q.: Limits of lightness identification for real objects under natural viewing conditions. *J. Vis.* **4**, 779–797 (2004)
9. Roorda, A., Williams, D.R.: The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522 (1999)
10. Smithson, H.E.: Sensory, computational and cognitive components of human colour constancy. *Philos. Trans. Roy. Soc. Lond. Ser. B Biol. Sci.* **360**, 1329–1346 (2005)
11. Smithson, H.E., Zaidi, Q.: Colour constancy in context: roles for local adaptation and levels of reference. *J. Vis.* **4**, 693–710 (2004)
12. Speigle, J.M., Brainard, D.H.: Is color constancy task independent? In *The fourth color imaging conference: color science, systems and applications* (pp. 167–172). Society for Imaging Science and Technology (1996)
13. Troost, J.M., de Weert, C.M.: Naming versus matching in color constancy. *Percept. Psychophys.* **50**, 591–602 (1991)
14. Witzel, C.: Misconceptions about colour categories. *Rev. Philos. Psychol.* **10**, 499–540 (2019)

Color Contrast

Lois Swirnof¹, Nilgün Olguntürk² and Gertrud Olsson³

¹The Cooper Union for the Advancement of Science and Art, New York, NY, USA

²Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

³Academy of Design and Crafts at Steneby, University of Gothenburg, Gothenburg, Sweden

Synonyms

Chromatic contrast

Definition

Color contrast describes the perceptual effects of colors' adjacency in contexts, whether they occur and are observed in two- or three-dimensional space. It is the relationship between the color of a stimulus and that of its immediate surround.

Overview

The concept of color contrast is concerned with the perception of color itself and is based upon the idea that the eye – the visual system – evaluates, a function of thinking. Light and color are therefore not merely surface phenomena but are intrinsic to human vision, to sight and insight. Light and color contrasts are primary attributes of vision. They signal the reactive eye to initiate and activate the complex visual system to respond to the visual world.

Fundamentally, the perceptions of the visual world, in its enormous diversity – the environment, man-made, or natural; people; places; animals; and vegetal life space itself – are distinguished and contextualized by the glorious capacity to *see*. The primary function of the visual system evolved to react to this diversity of features as complex contrasts of shapes and colors. These formal attributes originate and are embodied in the response of the human eye to gradients of light and its absence, shadow. Gibson links the concept of gradient to both formal and spatial effects, as well as other incremental changes [1]. In the context of color contrast, it is possible to use this concept to express increments of light and color. It is possible to think of this as a “quantification without number.” Color originates in the environment and in the visual system, as a complex and selective response to light.

Visual artists, as abstract or figurative painters, are experts in defining the images they create on two-dimensional surfaces by analyzing and expressing them as shapes and areas of light and color. All visual languages require this ability. Figurative painting depends upon the narrative orchestration of these components; for abstract works of art, shapes and colors, with their strong

associative attributes, suffice as expressive forms. Now, how does the visual system make sense in the first place of the complexity of the visual world? By responding to its gradients of light and colors as areas of contrast (Fig. 1).

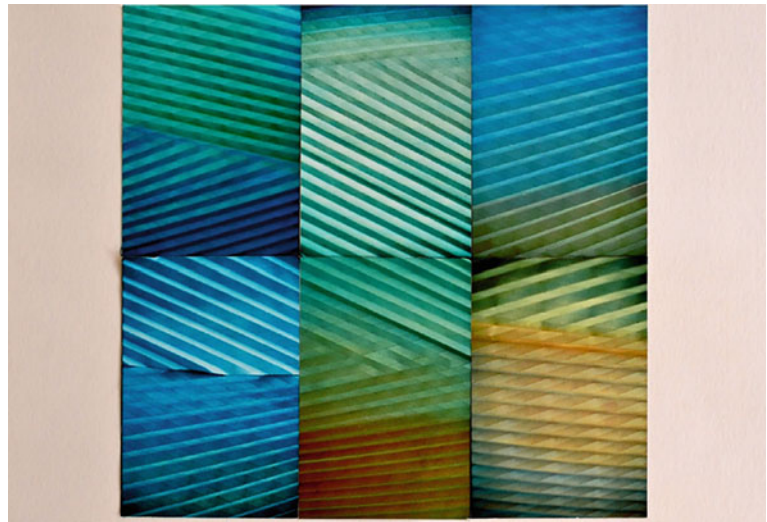
But color, as Josef Albers asserted, is the most changeable component in art. And the painter Delacroix boasted he could turn a mud color into gold. It is all a matter of context. Chevreul, the chemist, in the nineteenth century defined “simultaneous contrast” as the influence of colors in backgrounds to affect the appearance of colors within their boundaries [2]. Thus, for example, a

middle gray against a black background looks light, while that same gray against a light background appears dark. Lights and darks are particularly susceptible to change (Figs. 2 and 3).

Josef Albers, in the twentieth century, explored and exhaustively expanded the simultaneous contrast law, as well as the Bezold effect, the Liebmann effect, Fechner’s psychophysical laws, and Goethe’s perceptual concepts, by his and his students’ experiments and in his great work, *The Interaction of Color* [3]. In it, the empirical studies which entailed figure–ground relationships were explored far beyond Chevreul’s. To the

Color Contrast,

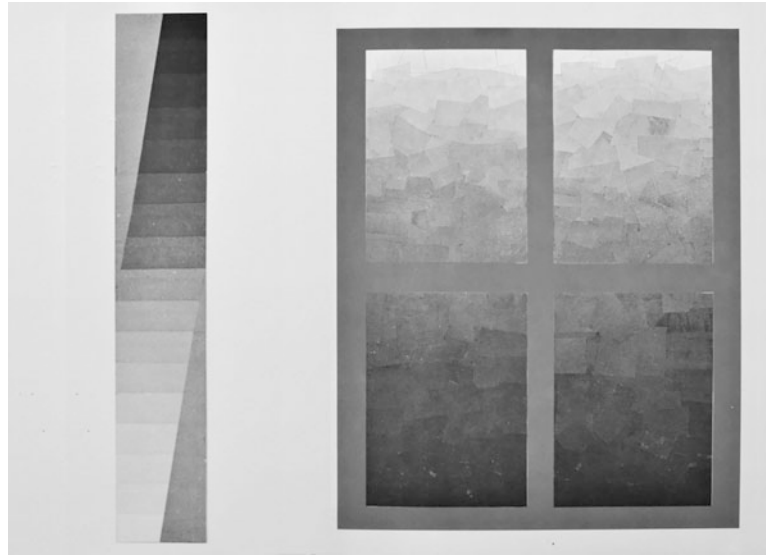
Fig. 1 Lois Swirnoff:
Santa Fe Rhythm, acrylic on
folded paper



Color Contrast,

Fig. 2 The same gray
appears as two



Color Contrast,**Fig. 3** Scales of gray

extent that contrasts between color complements, like a yellow and a violet, against opposing backgrounds, could appear to be similar (Figs. 4, 5, 6, and 7).

Underlying Albers' discoveries is the function of the visual system – the afterimage. The stimulus of a bright red area, after 30–60 s of exposure to the eye, will appear bright green. Just as staring at the bright light of a window in a darkened room appears as a dark window, when the eyes are diverted or shut. Another was the contrast boundary, the thin line at the edge of a color figure which appears against a contrasting background, initiating the reaction of a color change. Application of gradients of colors as linear planes can cause mixtures between adjacent color boundaries to appear as an additive third, i.e., a combined effect which regenerates the appearance of light. Changing the quantity of this gradient between two colors in a painting can cause many additive mixtures to be created. The work of Julian Stanczak and Richard Anuskiewicz, former graduate students of Albers, explore these phenomena and created the movement of Op Art in the 1960s in the USA (Fig. 8). The researchers Leo Hurvich and Dorothea Jameson describe this phenomenon in their article “From contrast to assimilation; in art and in the eye” [4] and attribute it to an

opponent response system in human vision, beginning at the retina. They are proponents of the theories of Ewald Hering, whose book *Outlines of a Theory of the Light Sense* they translated into English [5].

Swirnoff's experimental work with color began with a question: What happens to color interactions lifted from the two-dimensional plane when they interact in space? Will contrast boundaries prevail in adjacent colors when they appear as sequences of planes in space, as they do on the 2D surface? Do the size and/or placement of volumes appear different when their surfaces reflect contrasting hues? How are clues to size and distance influenced by color? What is the influence of reds (long wavelength) to blues (short wavelengths) on the appearance of volumes in space?

Beginning with Gestalt principles of form and organization, size and placement, proximity, clustering, and grouping, Swirnoff applied primary surface colors (Color-Aid) to cubes or floating planes, differing in size and placement in a space frame, and observed their effects through a series of experiments. These findings are the subject of the paper titled “Spatial Aspects of Color,” a thesis written for her Master of Fine Arts at Yale University in 1956.

Color Contrast,**Fig. 4** Reverse ground:
each X is the same color

Since then, Swirnoff's experiments with her students of design, over more than three decades, have explored contrast effects spatially, from tabletop experiments to collaborations of environmental scale, published in her book *Dimensional Color* [6]. These experiments have resulted in the assertion that color can be defined as a nonlinear dimension.

A new phenomenon was discovered: color stereopsis (Figs. 9, 10, 11, 12, 13, and 14). This occurs when three interrelated colors, separated sequentially in a space/frame, are observed

frontally through a square opening. In this study, the contrast boundary between an intermediary color, seen as adjacent to two, one placed behind, the other in front, disappears. With the absence of the middle intermediary, the visual system prefers to see the two color planes, which are frontally aligned in the space frame, appear to rotate to a diagonal position, and as they appear to intersect, trigger a strong stereoptic effect, analogous to the experience of depth when two flat images of the same subject, seen through a "stereopticon" – differing slightly in their position – fuse.

Color Contrast,
Fig. 5 Reverse
 ground: disc



Color Contrast,
Fig. 6 Three colors appear
 as two



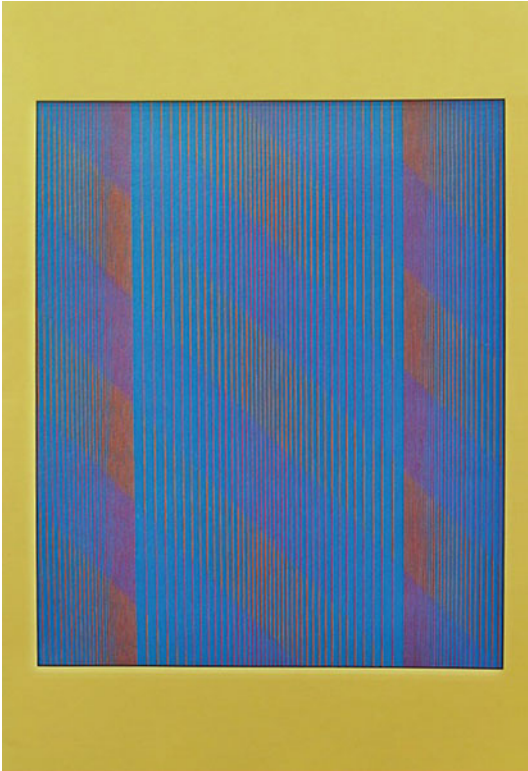
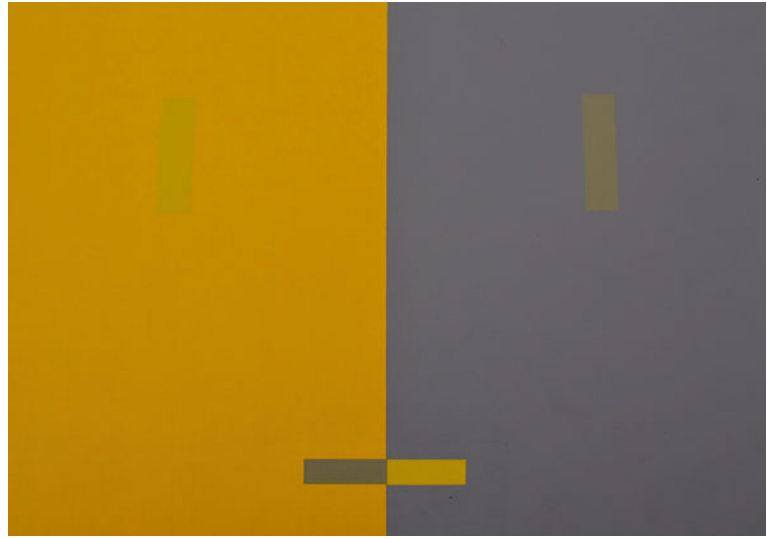
In a problem devised to test the degree of light and shadow that defines a square pyramid, models were placed against the walls of the teaching studio to observe the patterns of light and shadow which reflect from the four white triangles of its faces (Figs. 15 and 16). The challenge – to see what degrees of contrast produced the visual effect of a pyramid – entailed placing gray papers ranging from white to black against the model in reversed order, to cancel the pattern of light and dark, causing it to appear flat, without altitude. It was found that the degree of change in the contrast values perceived, with scales of gray papers,

between white to black was *geometric* – far greater than anticipated. Once found, the contrasting grays, applied to the four triangles in opposed order to their original appearance, matched. The contrast boundaries disappeared, and the model took the appearance of a flattened surface, a bisected figure or a pyramid of increased altitude, when rotated against the wall or in the hand. Thus, the response of the visual system to contrasts of light and dark depends upon *ratios* rather than measurable reflectances.

With a grant from the IIDA, Swirnoff was able to conduct a room-sized experiment, with students

Color Contrast,

Fig. 7 Complementary colors violet and yellow appear equal



Color Contrast, Fig. 8 Julian Stanczak: *Blue Squeeze*, acrylic on canvas

in the lighting studio of Parson's School of Design in New York. The issue was to test the effect of colored light, projected on interior surfaces. The

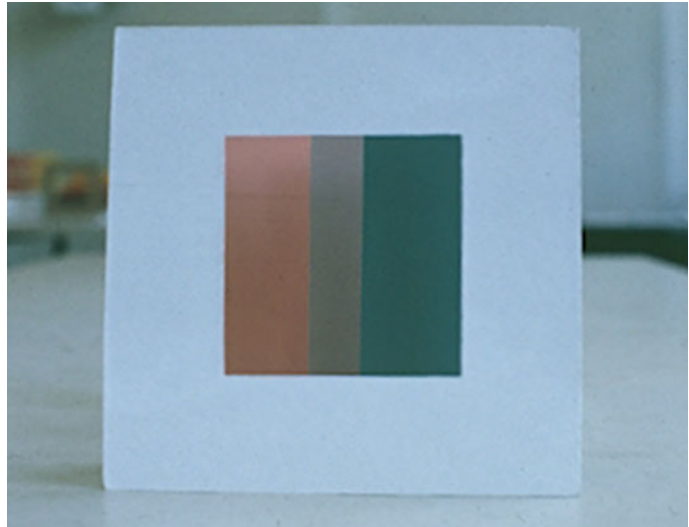
room's interior was constructed of four nine-foot squares, three comprised the walls – two lateral ones were connected by 90° to the third back wall – all three joined to the fourth square as the floor, an open cube. Using theatrical gels, colored light was projected onto these surfaces, and it was observed how they affected the spatial appearance of the room (Figs. 17 and 18).

Color generated by light greatly intensifies and enhances the experience of color contrasts. Both ends of the visual spectrum were tried (long wave, reds and yellows, then short wave, blues) sequentially, in separate experiments, and the differences in spatial experience by the contrasts each presented, as anticipated, were obvious. The “red room,” however, provided more phenomena [6].

Against the back wall adjacent to the lateral wall planes, illuminated with red, a neutralized yellow light was projected, to fill its square surface. At the 90° juncture color contrasts intensified; when observed over time, a neutralized yellow light was seen, reflected from the back wall, adjacent to the red lateral wall, change at their mutual boundaries, at first to a green edge and then followed by the green filling in the entire square! While the adjacent red wall increasingly appeared warmer, and at the contrast boundary, the 90° juncture with the green began to reflect a yellowish light.

Color Contrast,

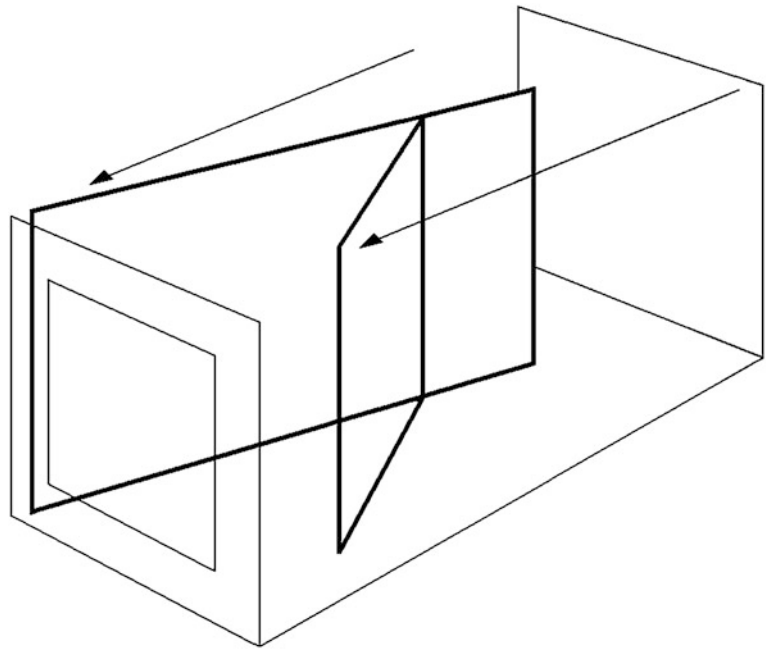
Fig. 9 Color stereopsis:
three colors fuse as two,
red/green



C

Color Contrast,

Fig. 10 Diagram: their
spatial appearance

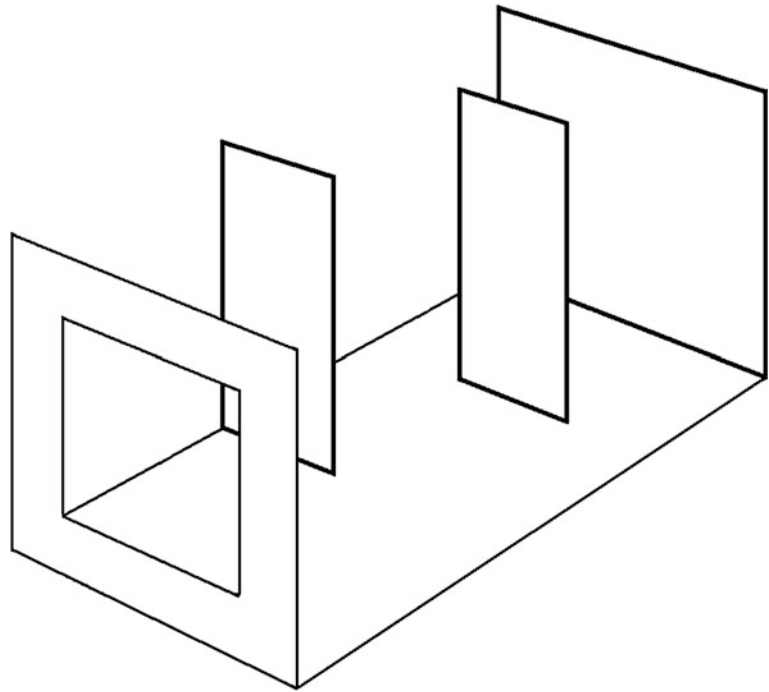
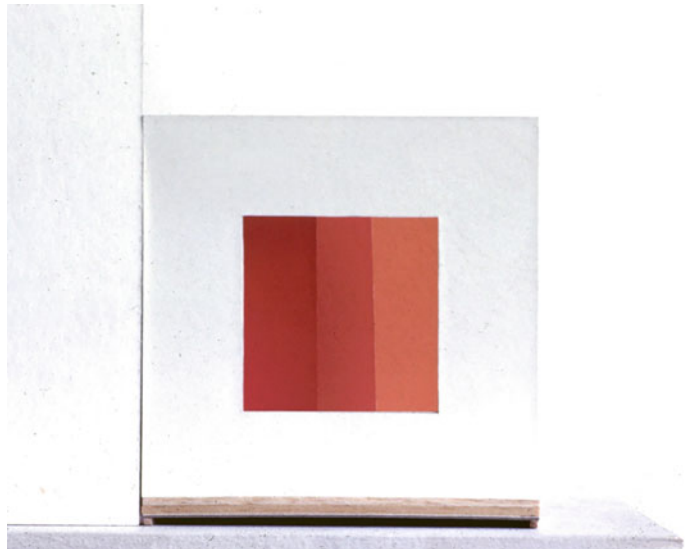


These changes were observed as they occurred over a period of about 8–10 min, when surprisingly, the green at its contrast boundary with the yellowing red changed to blue/violet.

There had been a complete change in the colors of the contrast boundaries, beginning originally as a yellow-projected light, which changed the entire plane to green, followed by a contrast boundary appearing gradually as blue/violet, now adjacent

to the red wall, whose contrast boundary appeared yellow – a complete reversal of contrast from the original stimuli of color.

Clearly, the intensity of the colored light provided a much greater stimulus to the visual system than that of a surface color. The color of surfaces induced to change by their adjacency on two-dimensional surfaces contrasts mutually (simultaneous contrast); once perceived, they

Color Contrast,**Fig. 11** Diagram: spatial placement of three planes**Color Contrast,****Fig. 12** Color stereopsis value sequence of three reds

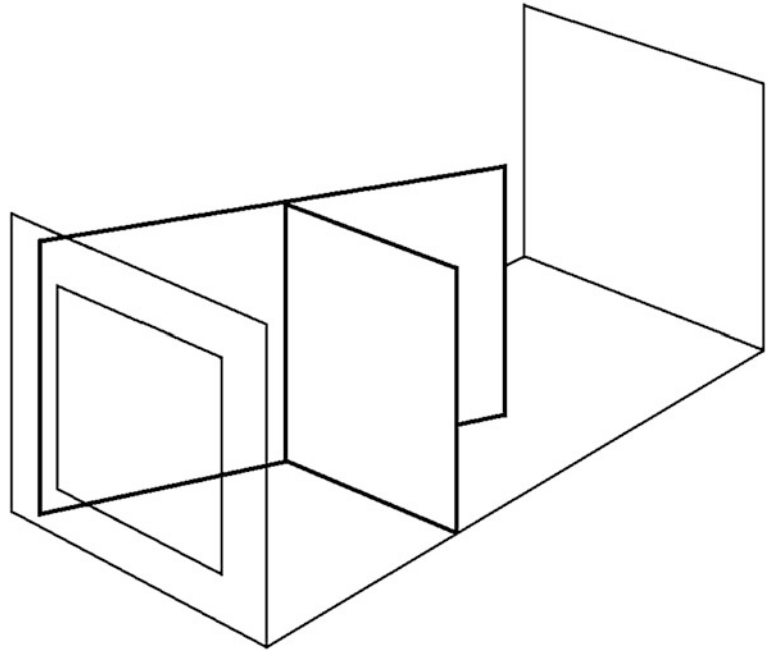
remain stable. This phenomenon, engendered by the intense light projected in space, seems to have extended to a sequence of perceptual changes which occurred after a much longer time – 8–10 min of observation. The experience all the observers had was *perceptual*. The effect eluded

all attempts to record it, by film or digitally. To account for these changes, elicited over a period of time, Swirnoff named the effect “sequential contrast.”

To conclude, the evidence raised by these experiments suggests that the issue of color

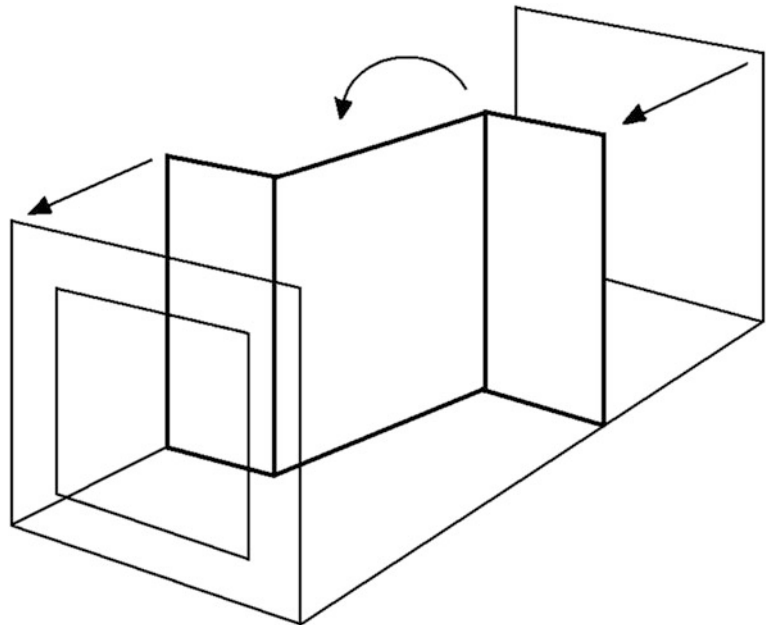
Color Contrast,

Fig. 13 Diagram: their spatial appearance



Color Contrast,

Fig. 14 Diagram: alternative appearance as zigzag



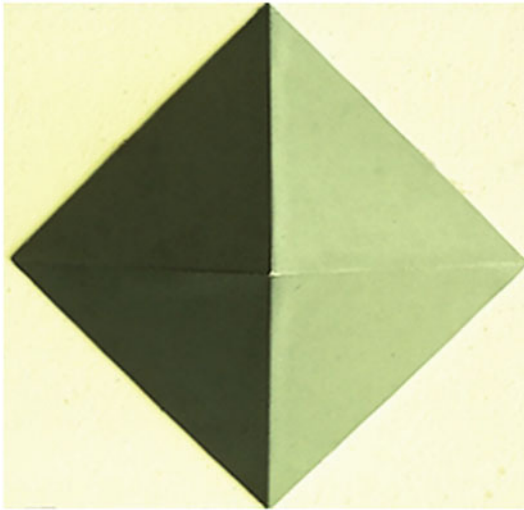
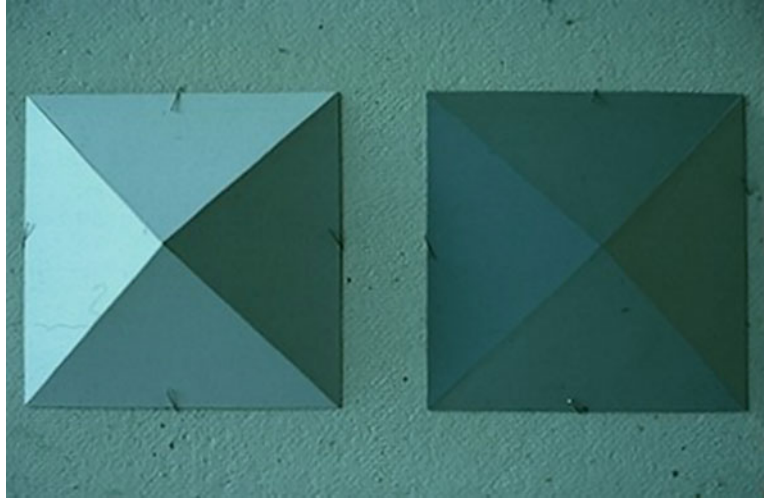
contrast is fundamental to and an essential part of the process of sight. Far beyond the purview of design science to explain, however, they offer a challenge now to neurology and brain science.

Effects in Perception

Contrast detection is the basic visual task from which all other visual behaviors are derived. The human visual system gives virtually no useful information unless there is a contrast in the retina

Color Contrast,

Fig. 15 Pyramids: *left*, the white model; *right*, model flattened by grays



Color Contrast, Fig. 16 Pyramid rotated 90° appears bisected

(thus, also in the environment that is being viewed). A small object, or a patch, can only be seen on a larger one if the two differ in color. These differences are known as contrast [7].

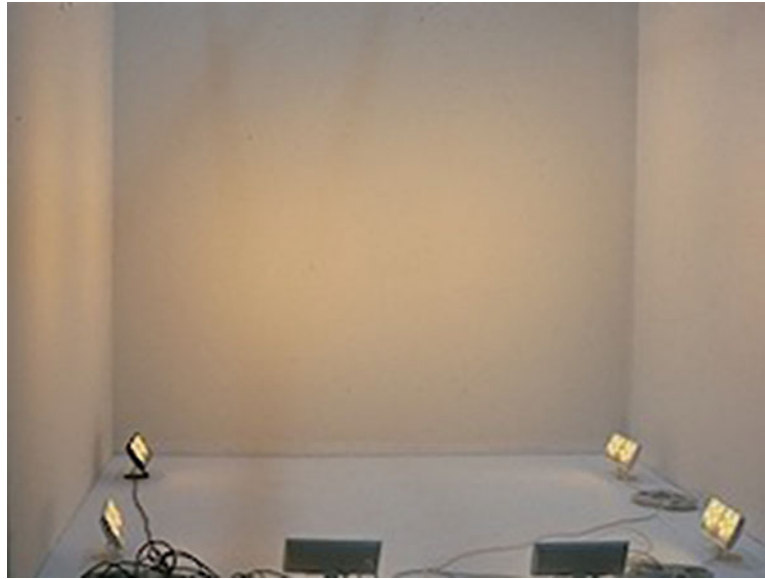
Sensation of color, or interpretation of color in the brain, is not only affected by adjacent areas of the stimulus, but also by the light under which the stimulus is seen. Two examples of this are the Helson-Judd effect and the Bezold-Brücke hue shift. The Helson-Judd effect is the tendency of lighter achromatic surfaces to take on the hue of the illuminant under which they are viewed and

darkers achromatic surfaces to take on the complementary hue. The Bezold-Brücke hue shift is a shift in the apparent color of a stimulus toward yellow or blue with the increasing intensity of light. If a pair of long-wavelength lights differing only in intensity is compared, the higher intensity stimulus will look more yellow and less red than the lower intensity light. For shorter wavelengths, higher-intensity lights look more blue and less green than lower-intensity lights [7].

Contrast perception also causes visual effects that lead to variance in color sensation. This may be caused by either the psychophysics of the eye or by the interpretation of the brain. These effects are: successive contrast, simultaneous contrast, edge contrast, and assimilation (reversed contrast).

Successive contrast is the visual effect which occurs when eyes are fixed on a colored patch for a sufficient period of time and then moved on to another patch of a different color. It is likely that the image of the first patch will be perceived upon the image of the second with its afterimage [7]. An afterimage is the visual effect that occurs after light stimulus has been removed. In the case of successive contrast, afterimage complementary color of the initially viewed image will be imposed on the lately viewed image.

Simultaneous contrast is the visual effect which occurs when two different color patches are viewed together, where both will exhibit

Color Contrast,**Fig. 17** The room illuminated by LED**Color Contrast, Fig. 18** The room's back wall changing to green

changes of appearance. Simultaneous contrast is affected by the distance between two colors. If two different color patches of equal size are placed side by side, both will exhibit changes in color appearance. When the patches are separated, the changes decrease and eventually diminish, as the distance between them is increased. Simultaneous contrast effects the hue changes by roughly superimposing the complementary color of the background on the foreground patch. This hue change is accompanied by changes in the perceived saturation depending on the afterimage complementary color produced. For example, if a blue-green

patch is viewed against a red background, the patch will be perceived as blue-green again (the afterimage of red being a blue-green), but with an increased saturation [8].

If two areas of uniform colors having the same hue but slightly different luminance factor are viewed, adjacent to the boundary between the two areas, there is a relative enhancement of lightness of the lighter area and a corresponding darkening of the adjoining area [8]. This is called edge contrast. If a black line is drawn along the edge at which two areas join, the effect of edge contrast is lost. The edge contrast phenomenon is caused psychophysically by interactions among nerve cells in the retina. It is also referred to as the Mach band effect, the Mach contrast (named after the physicist Ernst Mach), or the border contrast [7].

Assimilation (reversed contrast) is the visual effect which occurs when two distinctly perceived colors seem to shift in appearance toward each other. An example would be strips of colors, namely, red, yellow, and blue. The red strips on the yellow background appear yellowish, and on the blue background, the same red strips appear bluish. Assimilation is also known as the Bezold spreading effect. Assimilation should not be confused with simultaneous contrast. In simultaneous contrast, a red area surrounded by a yellow area

would tend to look more bluish (not yellowish); surrounded by a blue area, it would tend to look more yellowish (not bluish) [8]. It is also important to introduce a relation between the viewing distance and the effect that will be perceived. As the viewing distance is increased, producing retinal images of finer detail would be more possible, and a transition from simultaneous contrast to assimilation will occur. Finally, when a distance is reached beyond which the pattern is distinguishable, the visual experience of pointillism takes over where colors put on a surface will be perceived as a spatial mixture (optical mixture) [7].

Sonia Delaunay's Simultaneous Contrast

This concluding section focuses on Sonia Delaunay's and the early modernism's application of simultaneous contrast. For Sonia, and also for Robert Delaunay, and the poets Apollinaire and Cendrars, simultaneous contrast was a tool used to express the new era's speed, velocity, and movement. In Paris, the Russian-French artist Sonia Delaunay (1885–1979) worked with strong contrasts of color in her early expressionist painting period around 1906. A few years later, she started to work with simultaneous contrast in a nonfigurative way in several projects: paintings, fashion, ballet costumes, textiles, prints, books, and interior designs. As early as 1912, she created one of the first painting techniques in an abstract form with colors applied to large areas with the aim of achieving an interaction of color contrast [9]. She picked up the concept of simultaneous contrast from Chevreul and called her style of painting *simultanée*. Principally, she worked with color contrast to add power and chromatic strength to the tints.

Sonia – and her husband, the painter Robert Delaunay – first observed Chevreul's law in nature. In Spain and Portugal, she writes: the diffusion of light is the purest. The light is so strong that the colors themselves become distinct and their hues become robust. No haze or tones of gray interfere and mix the colors; no achromatic

grayed effects appear. The quality of this light allowed the artists “to go even further than Chevreul in finding dissonances in colored light” [10]. She explains “dissonances” as “rapid vibrations, which provoked greater color exaltation by the juxtaposition of specific hot and cold colors” [10]. To create the color vibrations, the Delaunays began to divide the shades of colors into hot and cold. This meant working with complementary colors and, to an even greater degree, with *cold* and *warm* color contrasts. Sonia Delaunay states that colors “agitated by hot and cold dissonances provoke a stimulating response” to the viewer [10].

Simultaneous Contrasts in Patterns

Sonia Delaunay's patterns in pure colors and in new color relationships are varied and have abstract forms: arcs and circles, rectangles, and triangles. For her, the circle was a symbol of the sun and also of simultaneous action. Her patterns change in rhythmic movements corresponding to their movements in value (lightness) and hue. Dresses and costumes are also inspired by the natural movement of the body. She began from the four basic colors – red, yellow, green, and blue – and from black and white. The red hue is often mixed with yellow to form an orange-red. The green is mixed with yellow to form a hue similar to the color of a mimosa flower. The blue is a medium blue, and the yellow is strong. In addition, she uses gray [11].

Around 1912, standard fabrics usually had large flowers against black or strongly colored backgrounds [9]. Sonia Delaunay's simultaneous fabrics changed this custom; she manufactured the same elements as those used in her paintings. Her abstract patterns are simple and have clear motifs in composition and color (quite similar to some African patterns). The colors are often vivid, but they harmonize most agreeably. Characteristic of her textile design is that the forms and patterns may well appear geometric, but the color surfaces are “characterized by *rhythm*” [10]. Due to Sonia Delaunay's simultaneous placement of the colors,

they produce new and original effects “right before your eyes” [10]. They are thus responsive to the architecture of modern life, to the new active way of living. In connection with this textile creation, she began to work with the first “simultaneous automobile,” a *Citroën B12* (1925), painted in the colors of the rainbow. In this way, she was at the forefront in the showing of art outside of the salons.

Simultaneous Contrast in Costume and in Fashion

During the summer of 1913, Sonia Delaunay began to design simultaneous dresses. She made and mounted them in collages made of textile. These dresses caught a new wave in fashion corresponding with the latest popular dances of the time, foxtrot and tango. In her abstract forms of arcs, circles, rectangles, and triangles, she created a movement of color. But the forms and the contrasting colors also enhanced the natural movement of the forms of the body, matching with the rhythms of Latin music. In the dancehall the Bal Bullier in Montparnasse, one could see the action of the dancers united with the action of color and light. Sonia Delaunay was commissioned to design the costume for the ballet *Cléopâtre* (1918). Working for the theater, she could experiment with successive designs for lengths of fabric: textiles wrapped around the human form, the body set into action in dance, all visual movements of the costume. Cleopatra’s costume was built up of discs in pure colors decorated with sequins and pearls. The ballet established Sonia Delaunay’s name as an innovator in both costume and fashion [12]. In 1922, a textile manufacturer in Lyon, France, asked her for a set of fabric designs and promptly ordered 50 designs for silk. For the commission, she began to study color relations and introduced abstract geometrical designs in printed silk. The subject of textile studies refined her control of the interaction of colors [11]. A few years later, on Boulevard Malesherbes in Paris 1925, Sonia Delaunay opened her own shop, the *Boutique Simultanée*.

There she offered simultaneous design in the form of coats, dresses, handbags, and even interior furnishings.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Appearance](#)
- [Chevreul, Michel-Eugène](#)
- [Chromostereopsis](#)
- [Color Harmony](#)
- [Fechner, Gustav Theodor](#)
- [Mach Bands](#)
- [Optical Art](#)
- [Unique Hues](#)

References

1. Gibson, J.J.: *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston (1966)
2. Chevreul, M.-E.: *The Principles of Harmony and Contrast of Colors* (English translation by Faber Birren). Reinhold, New York (1967)
3. Albers, J.: *The Interaction of Color*. Yale University Press, New Haven (1963)
4. Hurvich, L.M., Jameson, D.: From contrast to assimilation; in art and in the eye. *Leonardo*, **8**, 125–131 (1975)
5. Hering, E.: *Outlines of a Theory of the Light Sense* (Translated by Leo M. Hurvich and Dorothea Jameson). Harvard University Press, Cambridge, MA (1964)
6. Swinmoff, L.: *Dimensional Color*. Birkhauser, Boston (1989). Second revised edition: W. W. Norton, New York (2003)
7. Camgöz, N.: *Effects of Hue, Saturation, and Brightness on Attention and Preference*. Dissertation. Bilkent University, Ankara (2000). Also available by UMI, Bell & Howell Co., Ann Arbor (2001)
8. Agoston, G.A.: *Color Theory and Its Application in Art and Design*. Springer, Berlin (1987)
9. Damase, J.: *Sonia Delaunay. Fashion and Fabrics*. Thames and Hudson, London (1991)
10. Delaunay, R., Delaunay, S.: *The New Art of Color. The Writings of Robert and Sonia Delaunay*. The Viking Press, New York (1978)
11. Olsson, G.: *The Visible and the Invisible: Color Contrast Phenomena in Space*. Axl Books, Stockholm (2009)
12. Delaunay, S.: *A Retrospective*. Albright-Knox Art Gallery, New York (1980)

Color Contrast Adaptation

Katherine E. M. Tregillus¹ and
Michael A. Webster²

¹Center for Cognitive Sciences, University of
Minnesota, Minneapolis, USA

²Department of Psychology, University of
Nevada, Reno, NV, USA

Synonyms

Colorfulness adaptation; Habituation; Variance
adaptation

Definition

Changes in sensitivity or appearance in response
to spatial or temporal variations in color

Concepts

All sensory systems continuously adjust sensitivity or adapt to match their responses to the current stimulus context. In vision, these adjustments occur at many levels along the visual pathway and adjust to most if not all of the properties encoded by the visual system, from simple image features to high-level representations like the attributes of a face [1]. Thus adaptation to color is a specific example of a very general and ubiquitous process affecting everything we see.

Most studies of color or “chromatic adaptation” have focused on the changes in sensitivity and perception resulting from exposure to a change in the average color of the stimulus [2]. Chromatic adaptation begins as early as the photoreceptors, which like a camera must adjust their limited response range so that it is centered on the current light level. Since the different cone classes are each light adapting independently, a light that differentially stimulates the cones will lead to a change in the relative sensitivity to different spectra. For example, a longwave (reddish) spectrum is a strong stimulus for the L cones while

a weak stimulus for the S cones. Thus in the presence of this light the L-cone sensitivity will be reduced while the S-cone sensitivity will be increased. Sensitivity changes that occur independently in the cones and vary inversely with the light level are known as von Kries adaptation. Adjustments of this kind tend to renormalize color percepts so that the cone signals are equated for the current spectrum. Thus after adapting to the longwave spectrum the responses in the L and S cones will become more similar and the initially red light will appear more gray. In turn, a light that appeared gray before the adaptation will appear bluish after adapting, as a negative afterimage of the red adapting color. This chromatic adaptation is thought to play an important role in color constancy, by adjusting sensitivity to discount changes in the color of objects resulting from changes in the illumination [3].

However, in addition to the mean, color vision also adapts to the range or variance of the color distribution. These adjustments are known as contrast adaptation [2]. Contrast adaptation is again a very general process and has been very widely studied in the case of spatial vision, where it is the range of luminance levels that is typically varied [4]. For example, exposure to a high contrast grating (alternate light and dark bars) results in a loss in contrast sensitivity. The threshold contrast difference for detecting the grating increases, while the apparent contrast of visible gratings decreases. These effects are selective for the spatial properties of the grating. That is, sensitivity is reduced for patterns that have a similar size (spatial frequency) or orientation as the adapting grating, while little affected for spatially distinct patterns. This also again leads to negative aftereffects in the appearance of other patterns. After adapting to a tilted grating, a vertical grating appears tilted in the opposite direction (the tilt aftereffect) and the perceived size or spatial frequency of gratings appears biased away from the adapting grating. Like the example above of the cones, these results are again consistent with gain changes in different mechanisms or visual “channels” sensitive to different spatial properties of the stimulus, each adjusting according to the level at which they are stimulated. Indeed, measurements

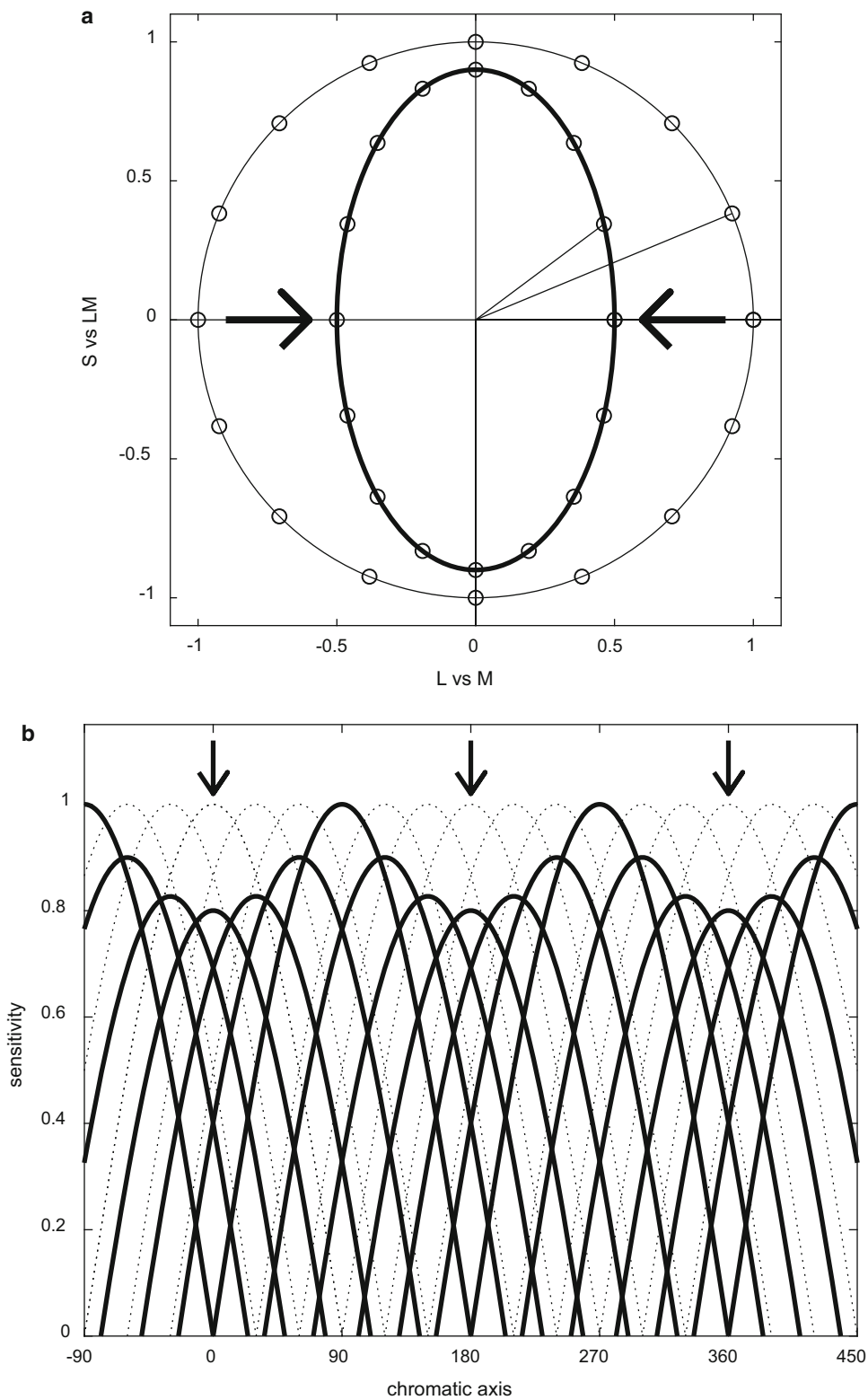
of the selectivity of the aftereffects of contrast adaptation were a primary tool in the study of spatial and temporal vision and in the development of channel models of visual coding [4]. For this reason, adaptation has often been described as “the psychologist’s electrode.” However, it should be noted that the aftereffects also depend on the network properties and interactions within the system and thus not simply on how each mechanism adapts in isolation [5].

History

Contrast adaptation in the context of color vision has received much less attention, and there are far fewer studies compared to chromatic (mean) adaptation. It is nevertheless an important and robust phenomenon with important functional implications. While some earlier studies examined adaptation to temporal color contrast (flicker), a study by Krauskopf, Williams, and Heeley [6] was seminal in exploiting color contrast adaptation to examine the color selectivity of the aftereffects. Their stimulus was a sinusoidal modulation of the color or luminance of a uniform field along different directions in color space. All had the same mean and thus produced the same state of chromatic adaptation. However the flicker reduced sensitivity to detecting a color change (an effect they called habituation), and these effects were again selective for the color of the adapting stimulus. For example, adapting to reddish-greenish flicker made it harder to detect these colors than other colors. These aftereffects could not be accounted for by response changes in the receptors, and instead pointed to desensitization in postreceptoral mechanisms that drew on different combinations of the cone signals. For instance, adapting to achromatic flicker strongly modulates the signals in all of the cones. This reduced sensitivity to luminance variations (in mechanisms that sum the cone inputs), but had little effect on chromatic variations (which are instead detected by mechanisms sensitive to the differences between the cone signals). The

pattern of selectivity revealed three primary adaptable mechanisms tuned to achromatic stimuli or to chromatic stimuli that varied the excitation of the L vs. M cones or the S cones versus the L and M cones. These were thus identified as the “cardinal directions” of color space and were subsequently shown to correspond to the color tuning of the primary pathways carrying information from the retina to the cortex [7]. The cardinal axes have since become a standard metric for designing and interpreting experiments in color vision, impacting thousands of studies. It is therefore noteworthy that it was through color contrast adaptation that they were first firmly established, though again surprisingly few studies have followed up on the properties and consequences of this adaptation.

Further analyses by Krauskopf and colleagues showed that there was also some selectivity of the adaptation for chromatic directions intermediate to the cardinal axes [8]. This finding established the presence of additional “higher order color mechanisms,” which have also strongly influenced modern approaches and theories about human color vision. Subsequently Webster and Mollon [9] explored how the adaptation to temporal color contrast affected color appearance. The adaptation produced large and selective changes in both saturation and hue. Specifically, adapting to a given axis in color space (e.g., a modulation along the LvsM cardinal axis) reduced perceived contrast (saturation) the most along the adapting axis, and biased the perceived hue of other chromatic directions away from the adapting axis and toward an orthogonal direction in color space (Fig. 1a). For example, the LvsM adaptation caused a previously pure blue or yellow hue to appear less like the LvsM (reddish-cyan) axis and more like the hues (purple or lime) of the S cardinal axis. Both the saturation and hue changes are consistent with a loss of sensitivity in the mechanisms tuned to the adapting mechanisms, though there is also evidence for a dissociation, since at high test contrasts, colors can exhibit a large hue bias despite small changes in their perceived contrast.



Color Contrast Adaptation, Fig. 1 (a) Effects of contrast adaptation on color appearance. The outer circle depicts a set of test stimuli with equal contrast but spanning different angles within the LvsM and SvsLM plane.

Key Principles

The adaptation-induced biases in hue are similar to spatial orientation aftereffects, but reflect tilt aftereffects in the perceived direction (hue) within color space. As with spatial orientation, the changes in color appearance from contrast adaptation can be selective for any arbitrary color direction, again implicating multiple mechanisms tuned to different directions in color space. However, unlike the tilt aftereffect, which at best produces a subtle bias in perceived orientation (about 2 deg), the biases in hue can be dramatic and up to 30 deg. or more. This difference is consistent with a broader tuning for hue mechanisms compared to orientation channels, at least at the level at which the adaptation is occurring. The common effects of adaptation across different stimulus dimensions point to common coding principles underlying the representations for different visual attributes [1, 4]. In particular, they suggest that the kinds of multiple-channel population codes that are widely assumed for form and motion perception also underlie the representation of color appearance (Fig. 1b). The representation implied by contrast adaptation is thus very different from the traditional two-channel (red vs. green and blue vs. yellow) model of opponent color theory, but is consistent with a wealth of behavioral and physiological evidence for multiple higher order color mechanisms.

In contrast to chromatic adaptation, the primary sites of the sensitivity changes in color and spatial contrast adaptation are thought to be cortical. This is suggested by the fact that the color aftereffects are also selective for spatial properties of the stimulus like orientation and that the adaptation shows some interocular transfer, properties that the neurons first become selective for in primary visual cortex. Cells earlier in the visual pathway show weak adaptation to chromatic

contrast, though the degree of adaptation observed depends on the approach to measure it. Neural correlates of contrast adaptation have also been found in the response changes in single cells or in the brain activity (BOLD responses) measured by fMRI. However, the actual relationships between different neural and perceptual measurements of contrast adaptation are complex and uncertain.

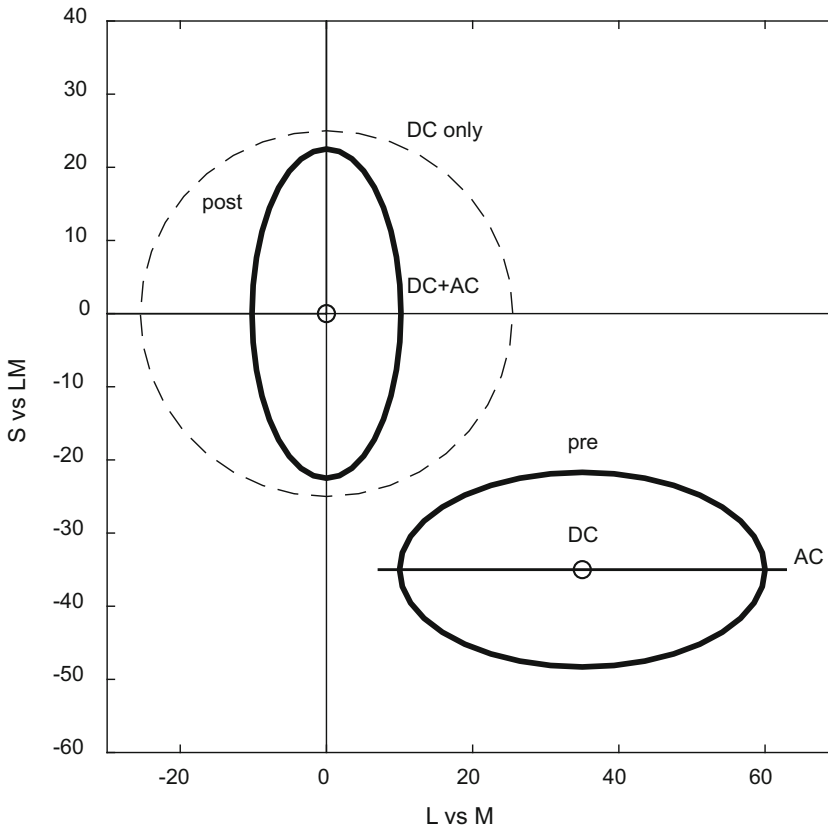
Thus in terms of their properties, chromatic adaptation and contrast adaptation reflect adjustments to two different statistics of the stimulus (the mean and the variance) arising at two different stages of the visual system (retinal and cortical) [2]. Consistent with this, the two types of adaptation produce separable effects on sensitivity and appearance (Fig. 2). Both types of adaptation also have a spatial analog. Just as a uniform color surround alters the hue of a central field (simultaneous color contrast), surrounding a field with a high-contrast background alters the perceived contrast, an effect which once more reflects very general processes of contrast normalization in sensory systems [10].

Phenomena

Both types of adaptation are important in natural viewing [1]. For example, the perceived contrast or colorfulness of natural or artificial images depends on the contrasts of the images that observers are adapted to. Color contrast in natural scenes is not random and instead tends to have a strong blue-yellow bias in part because of the dominant colors of earth and sky. The visual system is thus naturally exposed to stronger blue-yellow contrasts. Adaptation to this predicts the consistent finding that observers tend to be less sensitive to blue-yellow stimuli compared to other chromatic directions. These sensitivity

Color Contrast Adaptation, Fig. 1 (continued) Adaptation to the LvsM axis (arrows) selectively reduces sensitivity to the LvsM component of each test stimulus. This reduces perceived contrast (saturation) and also rotates the perceived hue angle away from the adapting axis. (b) Both

the contrast and hue changes are consistent with response changes in multiple mechanisms tuned to different axes in the chromatic plane. Adaptation to the LvsM axis (arrows) reduces the responses in mechanisms sensitive to this axis, thus biasing the distribution of responses across the channels



Color Contrast Adaptation, Fig. 2 Combined effects of chromatic adaptation and color contrast adaptation. “Pre” shows a set of colors centered on a yellowish mean (DC) chromaticity, while “post” plot the perceived colors after adaptation. Chromatic adaptation adjusts to the DC color so that it appears gray and shifts the set of colors so

that they appear centered on the origin (DC only). If observers are adapted to the same mean DC chromaticity but contrast is varied along the LvsM axis (AC,) then in addition to the mean color change there is a loss in perceived contrast along the LvsM axis (DC + AC)

differences are also evident in most uniform color spaces, which attempt to scale stimuli so that equal perceptual differences correspond to equal distances within the space. In these spaces the differences in cone signals are larger along the blue–yellow quadrant, implying weaker visual sensitivity. Ecosystems also vary systematically in their color contrasts and this suggests that individuals exposed to different environments (e.g., a forest or desert) may be held in different states of contrast adaptation. Moreover, contrast adaptation is also likely to be a factor in how the visual system adapts to the unnatural environments than humans increasingly occupy. An important case is the new generation of wide gamut lighting and displays. These increase the chromatic

contrasts that observers are exposed to, and this predicts that adaptation to this increase will consequently reduce their contrast sensitivity.

It is less well appreciated that both chromatic and contrast adaptation also adjust to variations within or between observers [1]. For example, it is in part because of these adjustments that color perception remains relatively stable as we age or as the world falls on different retinal locations [11]. The stability arises because in this case the visual system is recalibrating to match a world that itself remains stable. In the case of contrast an intriguing example is how the world might appear to a color deficient observer. Anomalous trichromats retain three kinds of cones but with only small differences in the spectral peaks of



Color Contrast Adaptation, Fig. 3 Monet's *Sunrise (Marine)*. The palette of the original left image was adapted to equate the mean and contrast to a color distribution

characteristic of natural outdoor scenes (after [14]). Digital image courtesy of Getty's Open Content Program

their two longer-wave pigments. This reduces the reddish-green contrasts signaled by these cones, yet the visual system could potentially adapt to the weakened contrast and thereby increase sensitivity so that these colors appear stronger. To mediate this stability, the mechanisms of adaptation must themselves remain stable with aging, and importantly, color contrast adaptation appears to remain robust across the adult lifespan, with potentially even larger adaptation effects in older populations.

The timescales of color contrast adaptation are not well characterized, and there is conflicting evidence as to whether or not the effect can build up with longer periods of adaptation. Strong after-effects of spatial contrast adaptation can occur with brief exposures, and at least some adjustments can occur within a few milliseconds, recalibrating perception with each eye movement. For other adjustments it is advantageous to average over longer times (e.g., to adjust to the average color or contrast of a scene rather than to each individual surface as it is fixated). There is also growing evidence for adaptation effects that unfold over hours, days or even months [12]. A classic example is the McCollough Effect, in which the color aftereffects are contingent on the spatial orientation of the adapting pattern [13]. These aftereffects may persist permanently until they are actively extinguished by adapting to

the complementary stimulus. However, it is not known how the McCollough Effect – or for that matter other types of long-term adaptation – are related to the short-term response changes measured in conventional studies of contrast adaptation.

Because the general ways that chromatic and contrast adaptation alter color perception are known in principle, it is possible to simulate how the world might look to observers adapted to different environments or different observers adapted to the same environment [14] (Fig. 3). This process of adapting the images rather than the observer also has the advantage that adaptation can be simulated for very long timescales that are impractical to study by changing the observer. Measuring visual performance for these adapted images can then reveal some of the perceptual consequences and advantages of contrast adaptation, by assessing what observers can see and do with images that have been adapted to better match their current sensitivity [15].

Function

The fact that adaptation is a hallmark property of all sensory processes suggests it confers essential advantages, and a wide variety of interrelated functions have been proposed [1]. One set derives

from information theory and emphasizes the role of adaptation in optimizing coding efficiency by matching the limited response range of neurons to the current range of stimuli. This also predicts the role of adaptation in normalizing or equating the responses across different mechanisms. This normalization is also closely related to ideas about how adaptation contributes to perceptual constancy. If the world or the observer changes, then adaptation implements a compensatory adjustment to discount the change and thus stabilize perception. A further related advantage is predictive coding, in which the sensory system adapts to remove the response to the absolute or expected stimulus level in order to preserve resources for signaling the relative differences or unexpected properties. In turn, this principle has been associated with the hypothesized effects of adaptation on heightening the perceptual salience of novel colors or spatial stimuli in the environment.

Cross-References

- [Adaptation](#)
- [Afterimage](#)
- [Color Information Processing in Higher Brain Areas](#)
- [Color Scene Statistics, Chromatic Scene Statistics](#)
- [Color Vision Testing](#)
- [Cortical](#)

References

1. Webster, M.A.: Visual adaptation. *Ann. Rev. Vis. Sci.* **1**, 547–567 (2015)
2. Webster, M.A.: Human colour perception and its adaptation. *Netw. Comput. Neural Syst.* **7**, 587–634 (1996)
3. Foster, D.H.: Color constancy. *Vis. Res.* **51**, 674–700 (2011)
4. Graham, N.V.: *Visual Pattern Analyzers*. Oxford University Press, Oxford (1989)
5. Solomon, S.G., Kohn, A.: Moving sensory adaptation beyond suppressive effects in single neurons. *Curr. Biol.* **24**, R1012–R1022 (2014)
6. Krauskopf, J., Williams, D.R., Heeley, D.W.: Cardinal directions of color space. *Vis. Res.* **22**, 1123–1131 (1982)
7. Lee, B.B.: Color coding in the primate visual pathway: a historical view. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **31**, A103–A112 (2014)
8. Krauskopf, J., Williams, D.R., Mandler, M.B., Brown, A.M.: Higher order color mechanisms. *Vis. Res.* **26**, 23–32 (1986)
9. Webster, M.A., Mollon, J.D.: Changes in colour appearance following post-receptoral adaptation. *Nature*. **349**, 235–238 (1991)
10. Carandini, M., Heeger, D.J.: Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2011)
11. Werner, J.S., Peterzell, D.H., Scheetz, A.J.: Light, vision, and aging. *Optom. Vis. Sci. Official Publication Am. Acad. Optom.* **67**, 214–229 (1990)
12. Tregillus, K.E., Engel, S.A.: Long-term adaptation to color. *Curr. Opin. Behav. Sci.* **30**, 116–121 (2019)
13. McCollough-Howard, C., Webster, M.A.: McCollough effect. *Scholarpedia*. **6**(2), 8175 (2011)
14. Webster, M.A., Tregillus, K.E.: Visualizing Visual Adaptation. *JoVE J. Vis. Exp.* **122**, e54038 (2017)
15. Webster, M.A.: Probing the functions of contextual modulation by adapting images rather than observers. *Vis. Res.* (2014)

Color Contrast Sensitivity (Rarely)

- [Chromatic Contrast Sensitivity](#)

Color Coordination

- [Color Combination](#)

Color Dictionaries and Corpora

Angela M. Brown
College of Optometry, Department of Optometry,
Ohio State University, Columbus, OH, USA

Definition

In the study of linguistics, a **corpus** is a data set of naturally occurring language (speech or writing)

that can be used to generate or test linguistic hypotheses. The study of color naming worldwide has been carried out using three types of data sets: (1) corpora of empirical color-naming data collected from native speakers of many languages; (2) scholarly data sets where the color terms are obtained from dictionaries, wordlists, and other secondary sources; and (3) philological data sets based on analysis of ancient texts.

History of Color Name Corpora and Scholarly Data Sets

In the middle of the nineteenth century, color-name data sets were primarily from philological analyses of ancient texts [1, 2]. Analyses of living languages soon followed, based on the reports of European missionaries and colonialists [3, 4]. In the twentieth century, influential data sets were elicited directly from native speakers [5], finally culminating in full-fledged empirical corpora of color terms elicited using physical color samples, reported by Paul Kay and his collaborators [6, 7]. Subsequently, scholarly data sets were published based on analyses of secondary sources [8, 9]. These data sets have been used to test specific hypotheses about the causes of variation in color naming across languages.

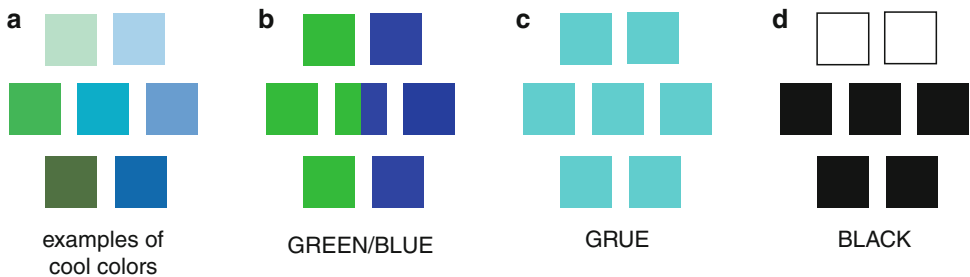
From the study of corpora and scholarly data sets, it has been known for over 150 years that languages differ in the number of color terms in common use. Particularly, languages differ greatly in how they name the cool colors that are called “blue” and “green” in English (Fig. 1).

Some languages, such as English, use a word *BLUE* that means only blue, in conjunction with a word *GREEN* that means only green. Other languages use a single term (here and elsewhere, “*GRUE*”) that means green or blue, and still other languages use a word (here, “*BLACK*”) that means both black and blue, to name the cool colors, in conjunction with *WHITE*, which names the light and warm colors.

Scholars in the nineteenth century established the two general explanations for this diversity of color terms across languages, which still guide much of the research on the topic today. The first explanation was that the people who spoke languages with few color terms had deficient color vision. This speculation was at first based on the philological analysis of extinct languages and arose in part because of general interest in the theory of evolution in the latter half of the nineteenth century. Proponents of this view speculated that humans and their color vision had evolved since ancient times. The second explanation was that people living at different times and in different cultures need to differentiate between different colors, so their languages have different numbers of color terms. Particularly, ancient languages lived in simpler times and consequently had fewer color terms in their lexicons.

The Color Deficiency Explanation

The earliest scholar to study this variability across languages was **William Ewart Gladstone**, prime minister of England over the latter half of the



Color Dictionaries and Corpora, Fig. 1 (a) Examples of cool colors; (b–d) false color coding of the corresponding color terms. The center sample of the

diagram is called *BLUE* by some informants and *GREEN* by others in (b), but it is called *GRUE* by all informants in (c), and it is called *BLACK* by all informants in (d)

nineteenth century and scholar of ancient Greek. Gladstone reported that Homer's epic poems used a "paucity" of color terms, mostly relating to dark and light, with a few instances of terms that may have corresponded to *YELLOW*, *RED*, *VIOLET*, and *INDIGO*, but not *GREEN* and not *BLUE* [1]. The German philologist **Lazarus Geiger** [2] reviewed evidence from even older sources: the Hindu Veda hymns of India, the Zend-Avesta books of the Parsees, and the Old Testament of the Bible, as well as ancient Greek and Roman sources. Geiger argued that color lexicons progressed over time from a *BLACK*-and-*WHITE* system to a *BLACK*-and-*RED* system (where *RED* was his term for white or warm colors), then differentiating *YELLOW*, then adding *GREEN*, then *BLUE*. "In the earliest mental productions that are preserved to us of the various peoples of the earth . . . notwithstanding a thousand obvious and often urgently pressing occasions that presented themselves, the colour blue is not mentioned at all. . . . Of the words that in any language that are used for blue, a smaller number originally signified green; the greater number in the earliest time signified black" [Ref. 2, pp. 49, 52].

Gladstone speculated that the Greek of the heroic age "had a less-evolved color sense that prevented him from seeing and distinguishing the many colors that modern people can see easily" [1] [p. 496]. Geiger came to a similar conclusion: "Were the organs of man's senses thousands of years ago in the same condition as now. . . ? [p. 60] The circumstance that the colour-terms originate according to a definite succession, and originate so everywhere, must have a common cause. This cause cannot consist in the primarily defective distinction merely. . . . [W]e must assume a gradually and regularly rising sensibility to impressions of colour."

The Cultural Explanation

Under the influence of the Darwinian thinking of the day, the English writer **Grant Allen** and the German ophthalmologist **Hugo Magnus** [23] thought that "primitive tribes" who lived in

modern times could provide information on the color naming and sensory color capabilities of ancient humans. Therefore, they sent questionnaires to Christian missionaries, explorers, and diplomats around the world, asking them about the color capabilities of the people they encountered and the color terms in their informants' native languages. Based on their responses, Allen wrote that "the colour-sense is, as a whole, absolutely identical throughout all branches of the human race" [Ref. 3, pp. 205] and afforded "a reasonable presumption in favour of a colour-sense in the earliest members of the human race." He therefore rejected the view, espoused by Gladstone and Geiger, that the reduced color vocabulary observed in many ancient and modern languages was due to a color vision defect. Magnus partly agreed, with qualification: "While some groups confirmed an awareness of colour, which rated in no way below that of the achievements of highly developed nations, others again gave proof of the lack of ability in identifying colours of middle- or short-wavelengths, and this was noted particularly in relationship to 'blue'" [Ref. 4, p. 145].

Based on the results of his surveys, Grant Allen proposed the second, cultural explanation of the diversity of color naming worldwide. "Words arise just in proportion to the necessity which exists for conveying their meaning. . . . Primitive man in his very earliest stage will have no colour terms whatsoever. . . . But when man comes to employ a pigment, the name of the pigment will easily glide into an adjectival sense. . . . [p. 259]. The further differentiation of the colour-vocabulary . . . is most developed among . . . dyers, drapers, milliners, and others who have to deal with coloured articles of clothing. . . ." "How then are we to explain the singular fact, which Mr. Gladstone undoubtedly succeeds in proving, that the Homeric ballads contain few actual colour-epithets? In the following manner, it seems to me. Language is at any time an index of the needs of intercommunication, not of the abstract perceptions, of those who use it."

Hugo Magnus became interested in the discrepancy between the excellent color awareness of many of the peoples in his survey and their

difficult color naming. His summary of how color terms co-occur in languages is reminiscent of the results of Gladstone and Geiger: "...while in some...communities the known terminology begins and ends with 'red', it stretches in other ones well beyond the 'yellow,' and with yet others, even beyond the 'green'." Magnus began his research under the influence of Gladstone and Geiger, but in the end he was also influenced by Grant Allen's work. Magnus concluded, "one might be tempted to formulate a... natural law of awareness – be that linguistically engendered or physiologically-anatomically conditioned as part of the natural growth of man."

Empirical Studies

The empirical tradition in the study of color terminology began with **W. H. R. Rivers**, a medical doctor and anthropologist, who traveled as an explorer to several parts of the world on behalf of the Royal Anthropological Institute. In his book *Reports of the Cambridge Anthropological Expedition to Torres Straits* [5], he compared the color vocabularies of three languages spoken by the Kiwai, Murray Islanders, and Western Tribes of the Torres Straits. "...As regards blue, the three languages may be taken as representatives of three stages in the evolution of a nomenclature for this colour. In Kiwai there is no word for blue; may blues are called names which mean black. ... while other blues are called by the same word which is used for green. In Murray Island there is no proper native term used for blue. Some of the natives, especially the older men, use [a native term], which means black, but the great majority use a term borrowed from English. ... The language of the Western Tribe of Torres Straits presents a more developed stage. ... [a native term] ... is used definitely for blue, but is also used for green. ... however, traces of the tendency to confuse blue and black still persist. ..." Rivers also reviewed scholarly evidence from ancient and contemporary sources, including his own work in Egypt and the Andaman Islands. All the empirical evidence he reviewed supported his view that the naming of blue is highly variable across

cultures: some call blue things *BLACK*, some call them *GREEN*, and some call them with a particular word for *BLUE*. He believed that *BLACK* was the most ancient term, *GREEN* was used in more developed societies, and *BLUE* was the most advanced color term.

In the twentieth century, the large-scale study of color naming across many languages was dormant until 1969, when **Brent Berlin** and **Paul Kay** published their monograph *Basic Color Terms: Their Universality and Evolution* [6]. Berlin and Kay collected a corpus of empirical color-naming data on 20 languages on individual speakers who lived in the San Francisco Bay area in the mid-1960s. They showed each subject an array of Munsell color samples and asked them to indicate the range of colors they assigned to each color term in his or her native language. Berlin and Kay augmented their corpus with scholarly data on the color lexicons of 78 additional languages, which were obtained from dictionaries and other scholarly sources. Berlin and Kay observed that all the color terms in all the color lexicons in their data set were drawn on a superset of only 11 universal **basic color terms**: *BLACK*, *WHITE*, *RED*, *YELLOW*, *GREEN*, *BLUE*, *BROWN*, *ORANGE*, *PINK*, *PURPLE*, and *GRAY*. They also observed that these color terms occurred together in only about seven different combinations. They speculated that these seven combinations of basic color terms represented seven ordered **stages** along an **evolutionary sequence** whereby the most primitive languages distinguish only *BLACK* and *WHITE*, and other color terms are added in a fixed sequence, until all 11 color terms are present. Berlin and Kay assigned each language in their data set to one of their seven stages of color term evolution. Their idea about the evolution of color terms was in line with the ideas advanced by Gladstone, Geiger, Allen, Magnus, and Rivers, although their explanation for the evolution of color terms was more in line with Allen's.

The methodology of Berlin and Kay and their theoretical interpretation of their data were criticized by others [e.g., Ref. 10]. Therefore, in the 1970s, Kay, Berlin, and their colleagues collected a new corpus of data on 110 world languages: the

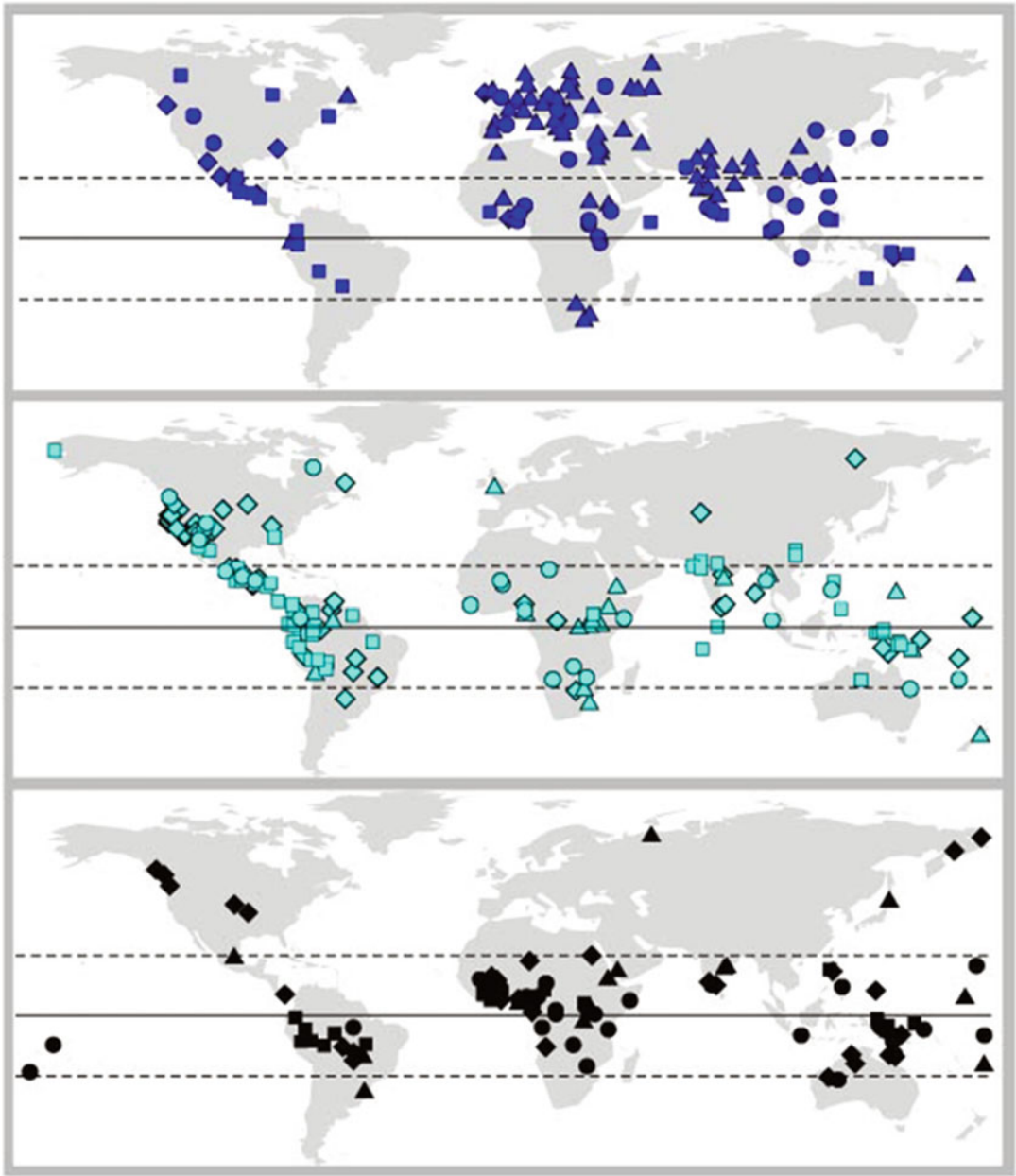
World Color Survey. The languages in the World Color Survey (WCS) were mostly unwritten and were spoken in traditional societies with limited contact with Western industrialized culture. The geographical distribution of the WCS languages was generally quite similar to the worldwide distribution of all living languages (www.ethnologue.com/show_map.asp?name=World&seq=10). The WCS data set was made up of empirical color-naming data provided in face-to-face interviews by about 24 speakers of each language. Each subject viewed 330 color samples, one at a time, and provided the color term they used in everyday life. Kay, Berlin, Maffi, Merrifield, and Cook published the *World Color Survey* [7], a book-length analysis of this corpus in which they identified each color term in each language with 1 of the 11 basic color terms of Berlin and Kay and updated their theory of color term evolution. They assigned each language to one of five stages, with two stages having three versions each, in their revised theory.

The **World Color Survey** corpus of color terms is available online and has been analyzed by Paul Kay and his colleagues [11, 12], who found evidence of universal color categories across the WCS languages. Independently, Lindsey and Brown [13] performed a cluster analysis of the color-naming patterns in the WCS corpus and discovered about eight distinct clusters of chromatic color terms, which, with the addition of the three achromatic terms *BLACK*, *WHITE*, and *GRAY*, corresponded approximately to the 11 basic color terms of Berlin and Kay. They further found that these color terms fell into about four color-naming systems (“motifs”) [14], which corresponded only loosely to the seven stages of Berlin and Kay or the five stages of the WCS. Similarly to all previous scholars and investigators, Lindsey and Brown found that the motifs differed most prominently in the color terms used to name cool colors that speakers of English call “blue.” In correspondence with the four motifs, some informants called blue samples *DARK*, some called them *GRUE*, some called them *BLUE*, and a few individuals called blue samples *GRAY* (a color term that was also used for middle-value neutral samples). Almost every World Color

Survey language revealed individual differences among its speakers, and at least three of the four motifs were represented among the speakers of most languages [see also Ref. 15]. Previous empirical and scholarly work that sought the color terms in each language as a whole, including the data sets in Fig. 2, could not reveal this prominent variation among individuals.

In the tradition of Allen, Kay and his colleagues [e.g., Ref. 16] argued that larger color lexicons, at a later stage along their evolutionary sequence, occur in technologically advanced, economically developed cultures, where the presence of colored artifacts and trade with other cultures requires a larger, more nuanced color vocabulary. Several authors [17, 18] have examined this hypothesis by comparing the number of terms that Berlin and Kay ascribed to each language to its level of development as published by [19].

Modern investigators have examined this diversity across languages in the naming of blue, which was common to the analyses of Gladstone, Geiger, Magnus, Berlin and Kay, and the WCS. Marc H. Bornstein [8] assembled a set of data on the presence or absence of *BLUE* in 145 languages from published sources and showed them on a world map. His analysis revealed a pronounced latitude effect: *BLACK* and *GRUE* languages tended to be spoken near the equator, and *BLUE* languages tended to be spoken at temperate latitudes (Fig. 2). Bornstein attributed this to the possible geographical variation in intraocular pigmentation, including the amount of melanin in the eye (the pigment epithelium and the iris) and the tint of the ocular media (the lens and macular pigment). Lindsey and Brown [20] reported that the yellow tint of the ocular lens, caused by the intense UVB in equatorial sunlight, could produce changes in color-naming behavior that were similar to those observed in *GRUE* languages. However, other work suggests that a causal link between the tint of the ocular lens and the naming of colors is at least partly modified by long-term chromatic adaptation [16, 21]. Brown and Lindsey [9] performed a geographical analysis of data on 118 ethnolinguistic groups for which both red-green color deficiency data (protan and deutan defects, not related to blue) and scholarly or



Color Dictionaries and Corpora, Fig. 2 The geographical distribution of languages that use BLUE/GREEN (top), GRUE (middle), or BLACK (bottom) to name the cool colors is known to be uneven worldwide. Maps show the localities of 371 living languages from color term corpora and scholarly data sets. Circles, 20 corpus languages and 77 living languages from scholarly sources from Berlin and Kay [6]; squares, 106 additional corpus languages from the World Color Survey [7], excluding two

already included in [6] and two that could not be typed; diamonds, 75 additional scholarly sources from Bornstein [8]; triangles, 93 additional scholarly sources from Brown and Lindsey [9]. Colonial languages (e.g., English, French, Spanish) are plotted where they were spoken in 1492 CE. Assignment of languages to BLUE/GREEN, GRUE, or BLACK types was made by the authors of the original data sets

dictionary data were available, also from published sources. The geographical results of Brown and Lindsey generally agreed with the results of Bornstein.

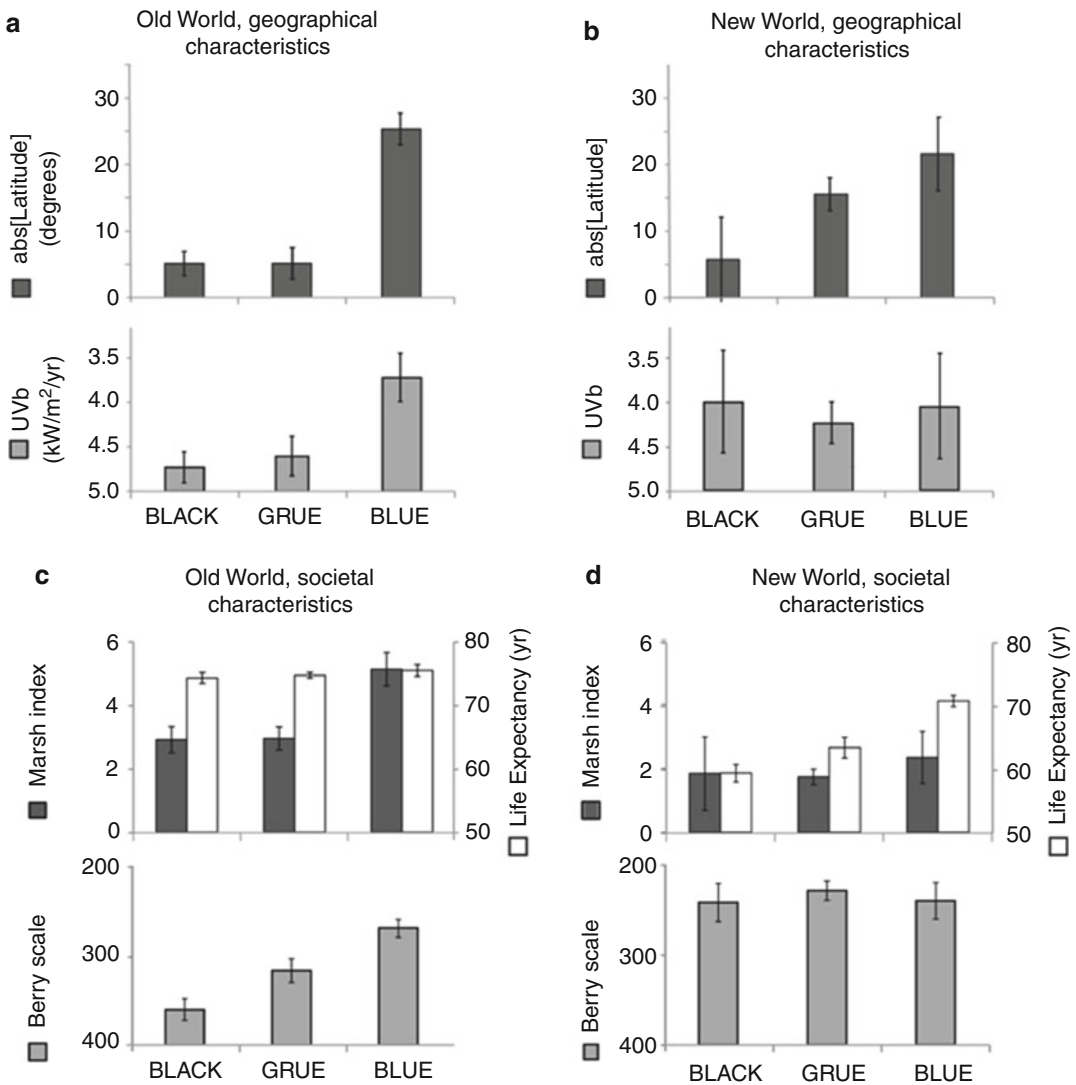
Color Naming Worldwide

There is still no single well-accepted explanation for the differences between languages in the use of *BLUE*, *GRUE*, and *BLACK*. Does *BLUE* vary across languages because of physiological differences among people, perhaps due to their different exposure to the sun, as Bornstein and Lindsey and Brown (and Gladstone and Geiger before them) suggested? If so, there might be a correlation between *BLUE* and the physical geography of the localities where these languages are spoken. Is the variation in *BLUE* due to the superior economic and cultural development of advanced nations, as Kay and his colleagues (and Allen and Magnus before them) suggested? If so, there might be a correlation between *BLUE* and the societal characteristics of the cultures where these languages are spoken. Or, is it a historical linguistic phenomenon, with the geographical patterns in the Old World being caused by the predominance of Indo-European languages in Europe? If so, then the non-Indo-European languages spoken in Europe should not show the predominance of *BLUE* observed in the Indo-European languages.

Figure 3, panels a, b, shows two physical geography characteristics of the individual cultures from Fig. 2, latitude and the annual dose of UVB. If the physiological hypothesis is correct, both of these graphs should correlate (with positive slope on the graph) with the use of *BLUE* (*BLACK* vs. *GRUE* vs. *BLUE*). Only latitude correlates with the use of *BLUE* in both the Old World and the New World. The idea that latitude has its effect through its effect on the annual dosage of UVB from sunlight has the difficulty that UVB dosage is correlated with *BLUE* in the Old World, but not the New World. The exposure of an individual person to UVB will be modulated by the amount of time he/she spends outdoors.

Three societal characteristics are shown in Fig. 3, panels c, d. These are Marsh's "Index of Societal Differentiation," which is a measure of the societal development of an individual culture [19]; life expectancy, which is a measure of the human development of nations; and Berry's "Technological Scale" of the economic development of nations [22]. Marsh's Index and Berry's Scale correlate with *BLUE* in the Old World but not the New World, whereas life expectancy was correlated with *BLUE* in the New World but not the Old World. These three societal effects are imperfect indicators of development. Marsh's Index is only available for the individual cultures where 166 of the languages in Fig. 2 are spoken. It is based on the per capita annual energy consumption and the fraction of males engaged in agriculture. Therefore, it will be correlated with latitude (it takes more energy to keep warm in Alaska than in Cameroon, and the short growing season makes agriculture in Siberia an unprofitable occupation). Life expectancy and Berry's Scale are available only for the nations within which the cultures are embedded. Life expectancy may show a ceiling effect in the Old World. Berry's Scale was the first principal component of an overall assessment of the economic advancement of nations, with contributions from transportation and trade, energy production and consumption, national product, communications, and urbanization. All of these societal characteristic correlates have the difficulty that none of them correlate with *BLUE* in both the Old World and the New World.

Languages that include *BLUE* greatly predominate Europe, the Mediterranean, and the Near East (Fig. 2, top panel). However, this is not entirely due to the predominance of Indo-European languages in that area. The data sets in Fig. 2 include 46 Old World languages that are spoken north of the Tropic of Cancer and west of the Ural Mountains. Of the 28 Indo-European languages and 18 non-Indo-European languages in this group, all but two use *BLUE*; one Indo-European language (Gaelic) uses *GRUE*, and one non-Indo-European language (Nenets) uses *BLACK*. Whatever is responsible for the predominance of *BLUE* in this geographical region, it is not entirely a question of linguistic heritage.



Color Dictionaries and Corpora, Fig. 3 Characteristics of the cultures or nations where the languages from Fig. 2 are spoken. Each bar indicates the value of the characteristic, averaged (± 1 s.e.m.) across cultures or nations having *BLACK*, *GRUE*, or *BLUE* as the word for blue in their languages' color lexicons. (a, b) Geographical characteristics of the localities where the 384 languages are spoken. Dark bars, absolute value of latitude, in degrees from the

equator; light gray bars, annual dose of UVB. (c, d) Societal characteristics. Dark bars, Marsh's "Index of Societal Differentiation," for 166 of the cultures shown in Fig. 2 [17, 19]. White bars, life expectancy; light gray bars, Berry's "Technological Scale" of economic development [22]. Life expectancy and Barry's Scale are shown for the nations where the languages are spoken

In spite of over 150 years of research, involving empirical corpora of color-naming data, scholarly data sets of color lexicons, and philological analyses of ancient texts, it is not well understood

why there is such great worldwide variation in the terms in the color lexicons of world languages. This topic continues to be the subject of much contemporary research.

Cross-References

- [Berlin and Kay Theory](#)
- [Evolution](#)
- [Unique Hues](#)
- [World Color Survey](#)

References

1. Gladstone, W.E.: *Studies on Homer and the Homeric Age*, vol. III. Oxford University Press, London (1858)
2. Geiger, L.: *Contributions to the History of the Development of the Human Race*. Trubner & Co., London (1880)
3. Allen, G.: *The Colour-Sense: Its Origin and Development: An Essay in Comparative Psychology*. Houghton, Boston (1879)
4. Saunders, B., Marth, I.-T.: *The Debate About Colour-Naming in 19th Century German Philology*. Leuven University Press, Leuven (2007)
5. Rivers, W.H.R.: Vision. In: Haddon, A.C. (ed.) *Reports on the Cambridge Anthropological Expedition of the Torres Straits*, pp. 1–132. Cambridge University Press, Cambridge (1901)
6. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley/Los Angeles (1969)
7. Kay, P., et al.: *The World Color Survey*. CSLI, Stanford (2009)
8. Bornstein, M.H.: Color vision and color naming: a psychophysiological hypothesis of cultural difference. *Psychol. Bull.* **80**(4), 257–285 (1973)
9. Brown, A.M., Lindsey, D.T.: Color and language: worldwide distribution of Daltonism and distinct words for “blue”. *Vis. Neurosci.* **21**, 409–412 (2004)
10. Hickerson, N.R.: Basic color terms: their universality and evolution by Brent Berlin; Paul Kay. *Int. J. Am. Linguist.* **37**(4), 257–270 (1971)
11. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
12. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. U. S. A.* **204**, 1436–1441 (2007)
13. Lindsey, D.T., Brown, A.M.: Universality of color names. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16608–16613 (2006)
14. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within-language diversity. *Proc. Natl. Acad. Sci. U. S. A.* **206**, 19785–19790 (2009)
15. Webster, M.A., Kay, P.: Variations in color naming within and across populations. *Behav. Brain Sci.* **28**(4), 512 (2005)
16. Hardy, J.L., et al.: Color naming, lens aging, and Grue: what the optics of the aging eye can teach us about color language. *Psychol. Sci.* **16**, 321–327 (2005)
17. Hays, D.G., Margolis, E., Naroll, R., Perkins, D.R.: Color term salience. *Am. Anthropol.* **74**, 1107–1121 (1972)
18. Ember, M.: Size of color lexicon: interaction of cultural and biological factors. *Am. Anthropol.* **80**(2), 364–367 (1978)
19. Marsh, R.M.: *Comparative Sociology: A Codification of Cross-Societal Analysis*. Harcourt, Brace & World, New York (1967)
20. Lindsey, D.T., Brown, A.M.: Color naming and the phototoxic effects of sunlight on the eye. *Psychol. Sci.* **13**(6), 506–512 (2002)
21. Delahunt, P.B., et al.: Long-term renormalization of chromatic mechanisms following cataract surgery. *Vis. Neurosci.* **21**, 301–307 (2004)
22. Berry, B.J.L.: An inductive approach to the regionalization of economic development. In: Ginsberg, N. (ed.) *Essays on Geography and Economic Development*, pp. 78–107. University of Chicago, Chicago (1960)
23. Schöntag, R., Schäfer-Prieß, B.: Color term research of Hugo Magnus. In: MacLaury, R.E., Paramei, G.V., Don Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modeling*, pp. 107–122. John Benjamins, Amsterdam (2007)

Color Directions

- [Color Trends](#)

Color Dynamics

- [Environmental Color Design](#)

Color Effects to Humans and the Environment

- [Functionality of Color](#)

Color Environment

- [Color Scene Statistics](#), [Chromatic Scene Statistics](#)

Color Experience

► [Color Phenomenology](#)

Color Fidelity Index

► [CIE Color Rendering Index](#)

Color Field Painting

Zena O'Connor
Design Research Associates, University of
Sydney, Sydney, NSW, Australia

Synonyms

[Hard-edge painting](#)

Definition

Color Field painting is an art movement that emerged in the mid twentieth century. Artists whose works can be categorised as Color Field chose to focus predominantly on the use of color in their works almost to the exclusion of other visual elements. Color Field paintings of the twentieth century are mostly works on canvas and some artists applied color in a formal, hard-edge manner while others chose to apply color in a more organic, freeform manner. The focus on color as demonstrated by the Color Field artists continues to influence contemporary artists; however, contemporary artists often explore color across a wider variety of media.

Overview

Of the three main currents in art that emerged in the twentieth century: Expressionism, Abstraction and Fantasy, Color Field painters were inspired by

the developments in Expressionism and Abstraction [1]. The focus of Expressionism was the artist's feelings and emotional responses via their conceptual content, subject matter and painterly technique; and the focus of Abstraction was a more conceptual approach to the partial or complete non-representational depiction of subject matter depicted with more formally structured painterly techniques.

Color Field painting emerged in the 1940s and painters such as Mark Rothko (1902–1970), Helen Frankenthaler (1928–2011) and Morris Louis (1912–1962) focused on expressing emotion through painting; while Clyfford Still (1904–1980), Barnett Newman (1905–1970) and Ellsworth Kelly (born 1923) applied a more formal, structured approach to their conceptual content and composition. The predominant focus that all Color Field painters shared was the use of color as the key conduit for conveying emotional or conceptual content.

Art critic Harold Rosenberg suggested that Abstract painting represented a new function whereby the canvas became an “arena in which to act. . . What was to go on the canvas was not a picture but an event” [2, p. 22]. From Rosenberg's perspective, “a painting is inseparable from the biography of the artist” [2, p. 23]. These views were countered at the time by art critic Clement Greenberg who preferred to focus on the formal qualities such as shape, color and line rather than the act of painting. Greenberg applauded the works of Still, Newman and Rothko with their primary focus on color and Greenberg marvelled that their paintings “exhale color with an enveloping effect that is enhanced by size” [3, p. 226].

The approach of the Color Field painters reflected the nineteenth century Impressionist painter's tradition of using ‘patches’ of color to capture a scene's ambience, convey a sense of movement plus capture a fleeting moment in time. This approach came about when *Baudelaire suggested* that painters should “evoke reality, not by detailing its forms, but by using a line or patch of tone to stimulate the spectator to recreate reality through the act of imagination” (Baudelaire, 1863, cited in [4, p. 9]).

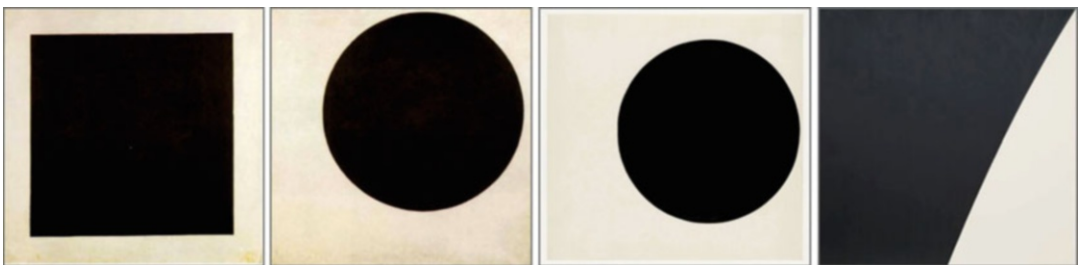
Color Field painters translated the Impressionists small-scale color ‘patches’ to much larger fields of color, encouraging their viewers to engage with their paintings on a more intimate but still static basis. Color was used to convey ambience as well as universal emotions such as *joie de vivre*, tension, tragedy and tranquillity. Rothko suggested that painting represented a “portal . . . into the vast recesses of the human psyche” and the role of color was to generate an emotional response: “I’m interested in eliciting basic human emotions – tragedy, ecstasy, doom and so on” (Rothko, cited in [5, p. 6]).

Using color as a form of communication or code, Color Field painters followed a long tradition. Vincent van Gogh (1853–1890) believed that color had the capacity to connect with the human condition: “I have tried to express the terrible passions of the human heart. . .” (Van Gogh cited in [6, p. 72]). In letters to his brother Theo, Van Gogh indicated that he attached emotional meanings to various hues and he clearly preferred strong, saturated color [7]. Similarly, Paul Gauguin (1848–1903) and the Symbolists believed that color could “act like words; that it [color] held an exact counterpart for every sensation, every nuance of feeling” [8, p. 129]. Likewise, Edvard Munch (1863–1944) used color and darker tones to help convey universal emotions such as anxiety, fear or sorrow in paintings like *The Dance of Life* (1899) and *Death in the Sick-room* (1895). Wassily Kandinsky (1866–1944) also believed that color as well as form had the capacity to communicate and he assigned certain connotations to specific colors; yellow for example represented warmth [9].

The conceptual and compositional approach of Suprematism and Russian Constructivism also had some influence on Color Field painting as evidenced by the patterns of similarity in the works of Kasimir Malevich (1879–1935) and El Lissitzky (1890–1941), and those of Rothko, Kelly and Newman. While there may be some conceptual or philosophic differences between the movements, there are strong similarities in the use of simplified geometric form as per Malevich’s *Black Square* (1915) and *Black Circle* (1915) and Kelly’s *Circle Form* (1961) and *White Curve* (1976) – see Fig. 1.

Mark Rothko (1903–1970) aimed to engage the viewer on a deeper, more personal level and “relied on large fields of color to produce solemn and elevated works” that had the power to convey something about the human condition [8, p. 314]. Eschewing other visual elements, Rothko suggested: “We may start with color”, which became the primary focus of his works (Rothko cited in [10, p. 65]).

Rothko said: “I’m interested only in expressing basic human emotions: tragedy, ecstasy, doom, and so on” [11]. The size of Rothko’s canvases, which often feature blocks of saturated color horizontally stacked, reveal his desire to create an intense experience for the viewer and he suggested that the ideal viewing distance for his works was a close arms-length distance from the canvas [5]. Rothko noted, “. . . I want to be very intimate and human. To paint a small picture is to paint yourself outside your experience, to look upon an experience as a stereopticon view or with a reducing glass. . . However, to paint the larger picture, you are in it. It isn’t something



Color Field Painting, Fig. 1 Patterns of similarity: Suprematism and Color Field painting

you command. . .” [8, p. 320]. Rothko’s paintings are a presence in themselves; so large and intensely colored that one is expected to feel their ‘Spiritual vibration’ [6, p. 72].

For Rothko, “A painting is not about experience. It *is* an experience”; an experience that transcends time and space conveying universal primal emotional experiences irrespective of gender, age and cultural experience (Rothko cited in [12, p. 57]). Rothko noted “My art is not abstract, it lives and breathes” (Rothko cited in [12, p. 50]). Key works by Rothko include *Untitled [Blue, Green, and Brown]* (1952); *Ochre and Red on Red* (1957); *Light Red Over Black* (1957); *Untitled* (1968); *Black Form No. 8* (1964), as per Fig. 2.

Barnett Newman (1905–1970) used the color palette of the De Stijl movement and the Bauhaus in a completely different way: “Why give in to these purists and formalists who have put a mortgage on red, yellow and blue, transforming these colors into an idea that destroys them as colors?” (Newman cited in [10, p. 27]).

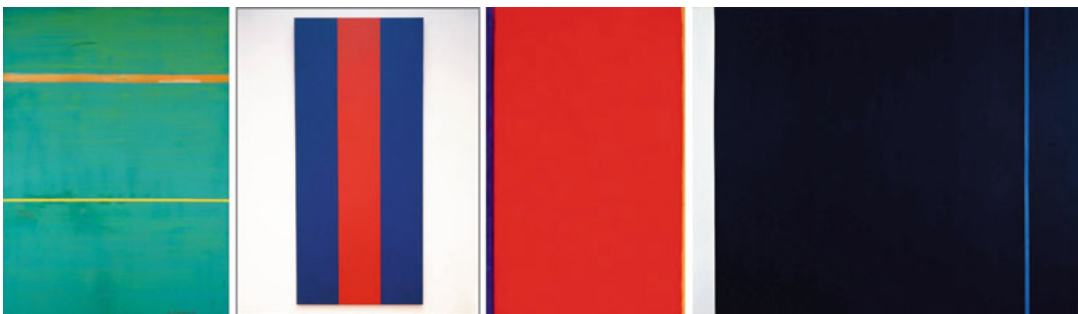
Newman commented that “The central issue of painting is the subject matter. . .my subject is anti-anecdotal” whereby a painting is more self-sufficient and independent with color and shape standing alone and prominent without reference to anything else [13].

Newman’s paintings often feature vertical lines that serve as “an act of division and creation” [between one color plane or reality and another]. . .the “zip” has become the single most dramatic event in the composition” [14, p. 77]. Key paintings by Newman include *Dionysius* (1949); *Voice of Fire* (1967); *Who’s Afraid of Red, Yellow and Blue?* (1966) and *Midnight Blue* (1970), as per Fig. 3.

Helen Frankenthaler’s (1928–2011) work became synonymous with a freer, more sensuous approach to Color Field painting [15]. Frankenthaler’s works are large and she applied color in a technique known as ‘soak stain’, wherein oil painted was diluted with turpentine so that the color soaked into the canvas creating halos of color.



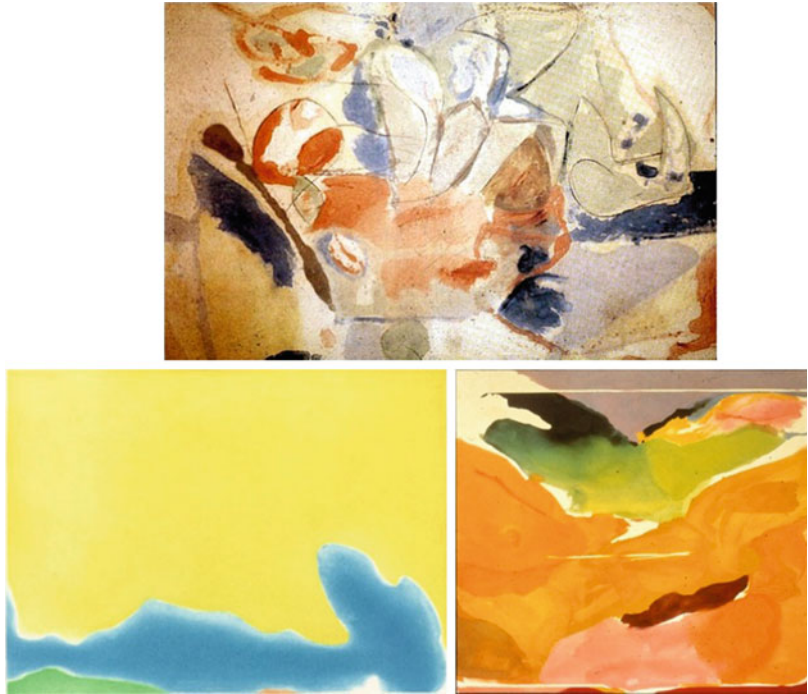
Color Field Painting, Fig. 2 Color Field paintings, Mark Rothko



Color Field Painting, Fig. 3 Color Field paintings, Barnett Newman

Color Field Painting,

Fig. 4 Color Field paintings, Helen Frankenthaler



The free-form and more sensuous nature of her paintings, prompted Hughes to suggest that “Landscape, imagined as Arcadia, remained the governing image in Frankenthaler’s work, and her titles often invoked the idea of Paradise or Eden [8, p. 154]. Key works by Frankenthaler include *Mountains and Sea* (1952); *Yellow Span* (1968) and *Nature Abhors a Vacuum* (1973), illustrated in Fig. 4.

The use of color by the New York-based Color Field painters influenced the West Coast painters as well as artists further afield, some of whom used new materials and techniques.

Karl Benjamin (1925–2012), whose work featured in the exhibition ‘Four Abstract Classicists’ (1959–1960), contributed to the US West Coast response to New York Abstract Expressionists. Benjamin’s work shared similarities with the Color Field painters and his works feature a sense of movement and vibrant optimism characteristic of the ‘California Cool’ aesthetic of California during the 1950s and early 1960s.

The use of color in Benjamin’s work in conjunction with the repetition of shape and line

creates works that are “fixed and stable” but in a state of “continuous flux” and convey a strong sense of movement and vitality [16]. Key works include *Black Pillars* (1957), #38 (1960), #7 (1966) and #7 (1986) illustrated in Fig. 5.

Lorser Feitelson (1898–1978), whose works also featured in the exhibition ‘Four Abstract Classicists’ (1959–1960), painted planes of color interspersed with a fluid use of line that imbued his layered abstract works with a strong sense of graceful movement, vitality and dynamism. Key works by Feitelson include *Hardedge Line Painting* (1963); *Untitled* (1965) and *Untitled* (1967) and *Coral and Blue Abstract* (1967), illustrated in Fig. 6.

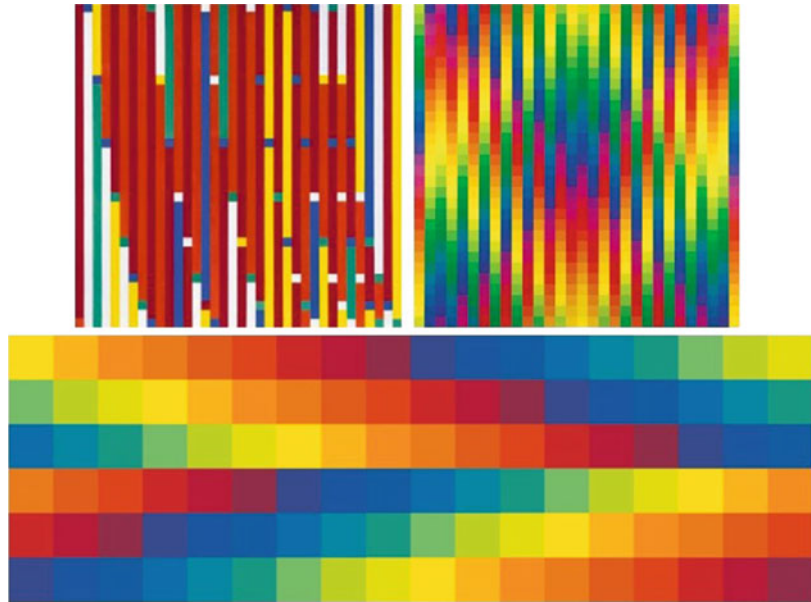
Richard Paul Lohse (1902–1988) created color investigations that “deploy the full range of spectral colors in endlessly inventive series” [10, p. 43]. Lohse painted individual colors in large grid formations to illustrate the notion that “the crowd contains the possibility of the individual” and also to highlight that color juxtaposition had the capacity to annul the “sovereignty of the [individual] color square” (Lohse cited in [10, p. 45]). Lohse’s works vary in scale from relatively small

Color Field Painting,
Fig. 5 Paintings, Karl
Benjamin



Color Field Painting,
Fig. 6 Paintings, Lorser
Feitelson



Color Field Painting,**Fig. 7** Paintings, Richard Paul Lohse

to very large, to enable an in-depth investigation of color and color juxtaposition [17]. Key works include *Progressive Reduction* (1942–1943); *Thirty vertical systematic color series in a yellow rhombic form* (1943, 1970) and *Thematic series in 18 colors* (1982) – see Fig. 7.

In the twenty-first century, artists have translated the idea of ‘patches’ of color and the approach taken by Color Field painters into a variety of new and different formats and mediums, often incorporating light, automated movement as well as the mechanics of human perception.

Gerhard Richter (born 1932) adopted the grid format of the early Modernists and imbued this with an abundance of color that could be grouped at random, as per the series of large-format paintings: *256 Colors* and *4900 colors* [10]. Richter’s random grouping of color mimics the complexity of color in nature and he painted several versions of *256 Colors* (1974) plus re-colored them in the 1980s. The work *4900 Colors* (2007) comprises 196 equal-size panels each containing 25 squares of color, as depicted in Fig. 8.

Australian Robert Owen (born 1937) also prefers depicting color via a grid format as it is “capable of providing infinite variations” and suggests that his paintings are “about levels of feelings, orders of sensation, shifting sequences of,

time and rhythm” [18, p. 9]. In *Cadence* (2003), Owen has depicted the range of his emotions measured using color tabulation. Responding to a competition for a public work by the Bureau of Meteorology, Owen said: “I thought, if they can measure atmosphere, I must be able to measure emotions. So, using a color tabulation, I intuitively picked how I felt every half hour during the day” [19]. The juxtaposition of color and constant variation triggers a strong sense of movement across the work. *Sunrise #3* (2005), painted in-situ in the Museum of Contemporary Art, Sydney, explores the impact of color by taking the Impressionists notion of ‘patches’ of color and translating this into a much larger work (Fig. 9).

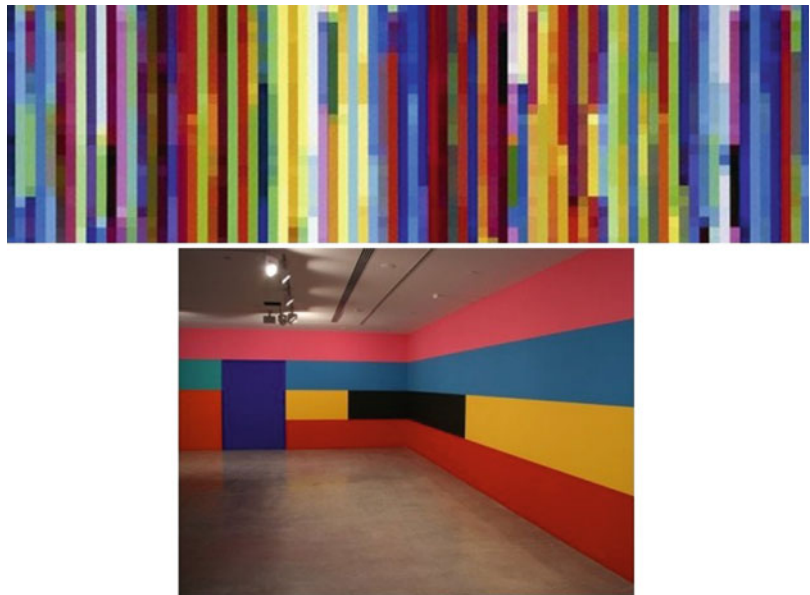
Los Angeles-based Dion Johnson (born 1975), whose works are reminiscent of the works of Benjamin and Feitelson, focuses on the interplay of color (*Gliger*, 2011). Colors converge and side-swipe each other, setting off color juxtapositions that add a sense of dynamism and vitality to Los Angeles-based Johnson’s large-scale canvases. Works such as *Velocity* (2012) and *Wild* (2012) feature hard-edge yet malleable color, as depicted in Fig. 10.

Rebecca Baumann (born 1983) has adopted the Color Field painters approach to color but added

Color Field Painting,
Fig. 8 Paintings, Gerhard
Richter



Color Field Painting,
Fig. 9 Paintings,
Robert Owen



mechanised movement. In *Automated Color Field* (2011), the automated movement brings ever-changing random juxtaposition to Baumann ‘patches’ of color. Baumann is interested in the

way ‘color is both universal and subjective’ with the capacity to ‘move people beyond cognitive and conscious thought’ (Baumann cited in [20]). Baumann acknowledges the changeable nature of

Color Field Painting,
Fig. 10 Paintings, Dion
 Johnson



Color Field Painting,
Fig. 11 Works by Rebecca
 Baumann (left) and
 Brendan Van Hek (right)



human response and her artworks feature “apparently random change(s) across the field, suggesting the flux in both our inner emotions

and the outside world” (Baumann cited in [21]), see Fig. 11 (left).

Brendan Van Hek (born 1968) uses found neon such as remnants of advertising signs and symbols

to create works like *Color Composition #3* (2013). In doing so, Van Hek subverts the original neon messages to create works that reflect the visual landscape and clutter of the twenty-first century: “*Color Composition # 3* started as an idea about landscape – a horizon line and the movement of forms along, below and above it (Fig. 11, right). What the work has developed into is one that describes a landscape of noises, activity, words, streaking taillights, constant action and movement – the city” (Van Hek cited in [22]).

Color Field Painting: The Influence on Design

The unique approach to color by the Color Field painters influenced art and applied design from the mid twentieth century through to contemporary practice. Before the advent of digital design, many designers were trained in colour mixing and application techniques influenced by the Bauhaus color curriculum and especially techniques devised by Josef Albers, Johannes Itten and

Color Field Painting,
Fig. 12 Film posters
designed by Saul Bass





Color Field Painting, Fig. 13 Poster and CD cover design inspired by Color Field painting

other former teachers and students of the Bauhaus. As a result, across all areas of applied design including graphic design, textile and fashion design, practitioners often used gouache paint to create designs, mixing colors as per the RYB color model.

Film posters designed by Saul Bass feature a hard edge, Color Field painting aesthetic: *The Man with the Golden Arm* (1955); *Anatomy of a Murder* by Saul Bass, (1959); *The Cardinal* (1963) and *The Human Factor* (1979), illustrated in Fig. 12.

Similarly, posters for *Before Sunrise* (1995), *The End of Summer* (2013) and *Rush* (2013) as well as recent ECM CD covers including Keith Jarrett's *Back Hand* (1974) and *Rio* (2011) are reminiscent of the Color Field paintings of Barnett

Newman and Ellsworth Kelly as well as Mark Rothko, as illustrated in Fig. 13.

Color Field painting also influenced textile and fashion design. The impact of this art movement continues with the fashion trend for color blocking that emerged in 2010 as well as the textile designs of Satu Maaranen (2011); Marimekko Sabluuna dress, Spring 2013 and the Gucci Sprint/Summer 2013 collection, illustrated in Fig. 14.

Color Field Painting: The Impact on Immersive Color Installations

Color Field painters influenced many subsequent artists, who create works that focus on color with



Color Field Painting, Fig. 14 Textile designs (Z. O'Connor) and fashion designs inspired by Color Field painting

the aim of connecting with the viewer and encouraging a deeper level of engagement often on a more intimate, immersive basis as per the works of Verner Panton, Peter Jones, Olafur Eliasson, Gabriel Dawe, Matthew Johnson and James Turrell. A common theme of these artists is exploration of the interface between color and human response in ways that leave the viewer enriched with a greater understanding about color.

Immersive works and projections are not viewed at arm's distance but often envelope the viewer, and elicit not only emotional response but trigger perceptual responses that form part of the work as a whole. While some of these works remain static, some works, such as those of Olafur Eliasson, involve a degree of 'radical subjectivity' on the part of the viewer, thereby making their perceptive subjectivity a component of the work [23].

Cross-References

- Art and Fashion Color Design
- Color Trends
- Impressionism
- Neon Lamp
- Palette

References

1. Janson, H.W.: *The History of Art*, revised and expanded by A.F. Janson. Thames & Hudson, New York (1995)
2. Rosenberg, H.: The American action painters. *Art News*. **51**(8), 22–23 (1952)
3. Greenberg, C.: *Art and Culture: Critical Essays*. Beacon Press, Boston (1961)
4. Green, P. (ed.): *Monet and Japan*. National Gallery of Australia, Canberra (2001)
5. Kelsey, A.: The color of transcendence. *Anamesa*. **3**(2), 3–14 (2005)
6. Bird, M.: *One Hundred Ideas That Changed Art*. Laurence King, London (2012)
7. Cabanne, P.: *Van Gogh*. Abrams, New York (1971)
8. Hughes, R.: *The Shock of the New: Art and Century of Change*. Thames & Hudson, London (1991)
9. Kandinsky, W.: *Über das Geistige in der Kunst*. R. Piper, Munich (1912). English transl. *Concerning the Spiritual in Art*. General Books, Memphis (2010)
10. Gage, J.: *Color in Art*. Thames & Hudson, London (2006)
11. MoMA: Mark Rothko. http://www.moma.org/collection/object.php?object_id=78485 (2013). Retrieved 29 Mar 2013
12. Baal-Teshuva, J.: *Mark Rothko: Pictures as Drama*. Taschen, Bonn (2003)
13. Gershman, R.: Barnett Newman. <http://www.theartstory.org/artist-newman-barnett.htm> (2012). Retrieved 2 Apr 2013
14. Adams, T.: He had to draw the line somewhere. *The Observer*, Sunday 22 September 2002, Features, p. 77. <http://www.guardian.co.uk/theobserver/2002/sep/22/features.review77>. Retrieved 2 Apr 2013
15. Glueck, G.: Helen Frankenthaler: abstract painter who shaped a movement dies at 83. *The New York Times*, 28 December 2011. <http://www.nytimes.com/2011/12/28/arts/helen-frankenthaler-abstract-painter-dies-at-83.html>. Retrieved 10 Mar 2013
16. Benjamin, K.: Karl Benjamin – Biography. <http://www.karlbenjamin.com/> (2013). Retrieved 5 Oct 2013
17. Richard Paul Lohse Foundation. Richard Paul Lohse: Biography. http://www.lohse.ch/bio_e.html (2012). Retrieved 10 Mar 2012
18. Museum of Contemporary Art, Sydney: Almanac – The gift on Ann Lewis AO. <https://www.mca.com.au/artists-works/exhibitions/almanac-the-gift-of-ann-lewis-ao/> (2009). Retrieved 19 Oct 2020

19. Art Gallery NSW: Robert Owen – Cadence. <http://www.artgallery.nsw.gov.au/collection/works/291.2004.a-e/> (2003). Retrieved 18 Nov 2010
20. Museum of Contemporary Art, Sydney: Rebecca Baumann – Automated Colour Field, 2012; Statement of significance by Anna Davis. <http://www.mca.com.au/collection/work/201120/> (2012). Retrieved 28 Mar 2013
21. Institute of Modern Art: Rebecca Baumann, Automated Monochrome Field 2011. <https://ima.org.au/exhibitions/rebecca-baumann-at-imaksubi/> (2012). Retrieved 19 Oct 2020
22. Lawrence Wilson Art Gallery, University of Western Australia: Luminous Flux exhibition <https://www.lwgallery.uwa.edu.au/exhibitions/past/2013/luminous-flux> (2013). Retrieved 19 Oct 2020
23. Grynstejn, M.E.: Take Your Time: Olafur Eliasson. Thames & Hudson, New York (2007)

Color from Motion

► Motion and Color Cognition

Color Harmonies for Fashion Design

► Art and Fashion Color Design

Color Harmony

Antal Nemcsics¹, Zena O'Connor² and Renata Pompas³

¹Budapest, Hungary

²Design Research Associates, Sydney, NSW, Australia

³Milan, Italy

Synonyms

[Balance](#); [Coherence](#); [Color accord](#)

Definition

A sense of accord and balance among colors in a visual composition or design, resulting in a positive affective response and/or cognitive judgment about color combination.

Color Harmony Theories

Many theories about color harmony have evolved since antiquity across different fields including art, design, psychology, and physics. Each of these has a different focus, and hence there is little consensus in terms of defining and describing the construct “color harmony.” Most theories about color harmony are highly prescriptive, and this is their weakness – they assume that responses to color are deterministic, uniform, universal, and fixed irrespective of context rather than idiosyncratic, less deterministic, and open to the influence of factors relating to culture and context.

The problem of color harmony has long been regarded as an esoteric matter for the artist. But since the advent of interdisciplinary research on colors, color harmony has become an object of scientific research giving rise to several theories. These theories emphasize different relations between the role of color harmony in environment and man, his culture, and his message and consider these as the exclusive laws of harmony. The scientifically established theories deal almost exclusively with the first level of the content of the concept of color harmony. They research the connections, which are mostly the same for all people. These relations express interactions between color perceptions, which can be described, in sum, in terms of relations between color perception parameters: hue, saturation, and lightness of harmonizing colors. These relations lend names to the harmony types such as complementary harmony, triadic harmony, scale of equal saturations, and others.

Color Harmony: A Twenty-First-Century Approach

Color harmony is a complex notion because human responses to color are both affective and cognitive, involving emotional response and judgment. Hence, our responses to color and the notion of color harmony are open to the influence of a range of different factors. These factors include individual differences (such as age, gender, personal preference, affective state, etc.) as well as cultural, subcultural, and socially based differences which give rise to conditioning and learned responses about color. In addition, context always has an influence on responses about color and notion of color harmony, and this concept is also influenced by temporal factors (such as changing trends) and perceptual factors (such as simultaneous contrast) which may impinge on human response to color. The following conceptual model illustrates this twenty-first-century approach to color harmony:

$$\text{Color harmony} = f(\text{color } 1, 2, 3 + n) * (ID + CE + CX + P + T)$$

wherein color harmony is a function (f) of the interaction between the color/s (col 1, 2, 3 + n) and the factors that influence positive aesthetic response to color which include the following: individual differences (ID) such as age, gender, personality, and affective state, cultural experiences (CE) and the influence of cultural and social conditioning, influence of the prevailing context (CX) of color(s) which may include setting and ambient lighting, intervening perceptual effects (P) such as simultaneous contrast, and, finally, the effects of time (T) in terms of prevailing social trends [1].

Harmony as a Balance Between Psychophysical Forces

According to the first theory on color harmony, the rules for producing color harmony experiences are determined by the mechanics of color vision. The trigger of the theory has been the phenomenon of successive contrast: longer

observation of green results in a red afterimage. Investigating this phenomenon, Rumford [2] has established that colors of successive color contrast mixed in an “appropriate proportion” in the Maxwell disk are resulting in a medium gray color. Goethe has built his theory of color harmony based on this observation. He thought the colors of successive contrast and simultaneous contrast are identical [3]. Therefore he stated that the “appropriate proportion” established by Rumford expresses the magnitude of color surfaces creating harmony experiences.

Chevreul, the French chemist, arrived at similar conclusions, stressing the role of complementarity in creating harmony, supporting his theory by successive contrast phenomena [4]. His message transmitted by Delacroix was translated by Seurat and Signac into the practice of painting [5–7]. Seurat himself experimented with harmonies, recapitulating his results by stating that harmony is a unity of opposites and similarities, a principle serving as the theoretical basis of pointillism.

Goethe’s theory has appeared with the mediation of Hölzel [8] in the works on the theory of art written by Kandinsky, Klee, Itten, Albers, and Moholy-Nagy [9–13]. Itten, in his book written for the students of the Bauhaus school, has fixed this proportion in relation to main colors [14]. He established that the appropriate mutual proportion of the colors yellow, orange, red, violet, blue, and green is as follows: 9:8:6:4:6.

Krawkow has demonstrated by experiments that the colors of successive contrast are not identical to the colors of simultaneous contrast [15]. This experimental result has refuted the conclusions of harmony theories originated from Rumford. Successive contrast is a physiological phenomenon, while the simultaneous color contrast is an aesthetic influence.

Harmony as Arrangement of Colors in Scales

It was observed by textile dyers and printers that mixing a color with white, gray, or black in different proportions led to very attractive, harmonic color complexes. It was one of the secrets of the

trade of painters to make each of their colors dull by admixing a color in different proportions, helping colors in the picture to form harmonic complexes. Both approaches arose from the recognition that colors with uniformly varying saturation or lightness, i.e., those forming a scale, appeared harmonic. This observation was the starting point of the second generation of theories on harmony.

The first scale of color harmony was constructed by Newton and published in his *Optics* in 1704 [16]. Dividing the spectrum into seven colors, he paralleled it to the musical scale. His idea was further developed by Hoffmann in his book published in 1786 [17], explaining color harmonies with the aid of acoustic analogies [18]. In 1810, the painter Runge suggested and attempted to develop a unified system of musical and color harmonies. These ideas misdirected the research on color harmonies for a long time since even their followers were concerned only with hue scales.

Ostwald [19] and Plochère [20] were the first to define the arrangement in scales by writing relationships between saturation and lightness. Scale members were described by Ostwald as additive color mixing and by Plochère as subtractive color mixing components. In his theory of harmonies, Ostwald pointed out that in order to find every possible harmony, possible orders in the color solid have to be found [21]. The simpler this order, the clearer and self-evident is the harmony. There are essentially two of these orders: those in the equivalent color circle and in the isochrome triangle. This latter statement expresses the dependence of harmony on the uniform variation of saturation and lightness. Ostwald's theory was progressive in that it connected the laws of color harmony with the relations between exactly measurable components. For the achromatic scale, the laws of harmony are coincident with experience. In other respects, however, experience fails to support his findings. The essential deficiency of his laws of harmony resides in his color system. Color points of his color system represent color perceptions related only by mathematically

definable quantitative variations of color mixing components. The interrelation between them comprises no perceptually equal or uniformly varying intervals.

Henri Pfeiffer, a painter starting from the Bauhaus school, later a graduate from the University of Cologne who became professor at the École d'Architecture in Paris, considered the creation of a color harmony system as his chef d'oeuvre. He started from the acoustic meaning of "harmony" – namely, that the three main tones *do* (*C*), *mi* (*E*), and *sol* (*G*) of a vibrating chord are proportional in a way that the chord length for *mi* (*E*) is the harmonic mean between the chord lengths for *do* (*C*) and *sol* (*G*). The algebraic generalization of this rule was applied for determining color harmonies. Referring to tests by Rosenstiehl [22] and Fechner [23], Pfeiffer established that lightness intervals of the logarithmic scale are in a harmonic relation and called this scale a harmonic scale. Then he deduced correlations between the logarithmic scale and the scale obtained by the golden section. In his book he described in detail his tests using Plateau and Maxwell disks to create harmonic scales. He classified his scales into two groups: isochromic or equal hue scales and isophanic or equal lightness scales. First he defined each kind of harmony and then presented its mode of construction by means of a revolving disk, followed by the analysis of the psychic character of harmony. Characteristic of his work are his chromatologic tables, with the aid of which he defined harmonizing color groups.

Harmonic scales can be developed not only according to laws of additive or subtractive color mixing but also by taking perceptually uniform intervals between colors. Colors of the Munsell color system constitute such perceptually uniform hue, saturation, and lightness scales [24, 25]. In his book, Munsell writes that a color scheme along a "path" is always harmonious.

Color Harmony as an Aesthetic Experience

According to the most recent theories of color harmony, the principles related to the establishment of color harmony experience could be

recognized only by statistical surveys based on a multitude of experiences. Their experiments are related to determined population each time. Ultimately they intended to create a system which may predict whether a given color composition will be judged by the members of a defined population as harmonic or not.

Moon and Spencer developed a model to quantify color harmony by statistical methods based on surveys [26, 27]. The main elements of Moon and Spencer's color harmony model are color intervals, area ratio, and aesthetic measurement. The aesthetical measure developed by Moon and Spencer, despite of its controversial results, provides a basic quantitative approximation of the problem in a field possessing so far only qualitative contemplations lacking verification. According to Moon and Spencer, the colors can be harmonious if the color difference is determined between the individual components. The definite color difference is defined as an interval. The magnitude of intervals defined by Moon and Spencer differs by hues. According to Moon and Spencer, a harmonic balance between color samples can be only reached if the scalar momentums related to an adaptation point in a linear space are equal to or are products of each other. The adaptation point is a point of the linear color space which corresponds to the adaptation state of the eye, normally it is medium gray.

More people have dealt with the definition of the magnitude of those intervals between members of a scale being judged as harmonic one. Dimmick [28] and Boring [29] determined the smallest of such intervals. They found that for an interval less than a certain value no harmony can any longer develop. Moon and Spencer found that these intervals were different for different hues. To examine and confirm this observation, Mori et al. [30] made experiments, and their findings agreed with those of earlier tests on lightness intervals by Katz, Gelb, and Granit [31–33].

In fine arts, intervals between colors in a picture have always been important and

characteristic of a given painter. Matisse [34] has always striven for uniform intervals in his color compositions. As Mondrian [35] wrote it in 1945, "For millennia, painting expressed proportionality in terms of color relations and form relations, achieving but recently the finding of proportionality itself."

The results of a line of experiments registering the judgments of more than 50,000 observers have been used by Nemcsics [36, 37] to create the color harmony index number system. The system denotes every existing couple of colors with a number between 0 and 100. This number expresses the magnitude of harmony of a color couple felt by different groups of people, i.e., how much is the harmony content of the color couple. The conclusions of the color harmony theory of Nemcsics are the following:

1. Between the Coloroid saturation and luminance values of color compositions felt harmonic, there is an identical number or logarithmically changing number of harmony intervals.
2. The members of compositions consisting of identical hues felt harmonic are located on straight lines of the actual axis section of the Coloroid color space [38].
3. The magnitude of harmony content of color compositions located on straight lines of the Coloroid color planes spaced at identical harmony intervals is felt different depending on the angle in relation to a straight line chosen as starting line being perpendicular to the gray axis.
4. The harmony content of hue pairs is higher than that of the others if they decline from each other in the Coloroid color space less than 10° , or decline between 30 and 40° , or decline between 130 and 140° , or their declination is near to 180° .
5. Both for men and women, it is valid that the intensity of the harmony experience of a composition is in synchronism with the preference of the colors in the composition bearing harmony.

Philosophical and Historical Overview

A great deal has been written about color harmony, from Chevreul [4], who was convinced that many different color hues and their harmony could be defined by means of the relationship between numbers and his color system as an instrument to find “harmony of analogy” and “harmony of contrasts” or from the numeric naming system of Munsell (1905) [24, 25], based on the criterion of perceptual uniformity and balance, passing through Bauhaus theories (1919–1933), until sophisticated contemporary three-dimensional color softwares.

However, in philosophical terms color harmony is a highly variable concept that is open to the influence of different including individual differences (such as age, gender, personality, affective state, and so on), cultural and subcultural differences, as well as contextual, perceptual, and temporal factors. In addition, from an ontological perspective, many early theorists championed an understanding of color harmony that was universal in nature and strictly deterministic, that is, color harmony was an effect that occurred for all people on a highly predictable basis irrespective of the context or situation. In the twenty-first century, theorists tend to take a more idiographic, stochastic view about color harmony. That is, the highly subjective nature of responses to color is acknowledged, and responses about color harmony are understood to occur on a more individual and less universal basis and are more likely to be probabilistic rather than deterministic and predictable [1, 39].

The art historian and theoretician Brandi (1906–1988) [40] writes that “the significance of words is not things, but the pre-conceptual scheme of things . . . , in other words, a summary of knowledge of things according to a given society”; the same is for the taste of color, connected with the geographical, historical, and social environment. This notion, which dates back to the ancient Greek philosophers, was also expressed by Hume (1711–1776): “Beauty in things exists

merely in the mind which contemplates them” [41].

Color Harmony as Historical Model

Since all that relates to matters of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – the definition of chromatic harmony periodically renews its model, creating new codes for new contents.

In olden times man sought perfection of his surrounds built on symmetry, as in the numeric unit of measurement of Greek temples, whose aesthetic quality comes about through internal, independent ordering of the whole. During the Renaissance, the architect Alberti (1404–1472) [42] defined beauty as “a harmony of all the parts . . . fitted together with such proportion and connection, that nothing could be added, diminished or altered, but for the worse”.

The different societies of the world have given rise to its own chromatic spectra, applying different criteria of harmony to the juxtaposition of colors, in a dynamic scheme, related to two essential properties: mimesis and ostentation.

Color Harmony as Mimesis

As in linguistics, there is onomatopoeic harmony, consisting in imitation of natural sounds by means of a phonic impression of the linguistic form, and in the animal kingdom, some creatures take on the chromatic features of the surrounding environment in order to camouflage themselves, so too in certain cultures, manufactured articles are modeled on the colors of the materials of the place in which they are introduced, thereby creating imitative harmony. Uniformity and similarity refer to something preexisting and therefore evoke aesthetic pleasure by reason of the fact that they are pleasing and reassuring.

Color Harmony as Ostentation

And yet, on the other hand, chromatic pleasantness is sought by enriching places with missing colors, so that full-bodied, vibrant, and rich colors stand out against the uniform backgrounds of

landscapes, acting, contrary to imitation, through differentiation. In both aesthetic models, mimesis and differentiation, there is harmony when there is unity in multiplicity, ordering the various parts into a coherent whole, when colors in a multi-chromatic background establish a dynamic and balanced relationship among dimension, distribution, saturation, whiteness, and blackness.

Color Harmony in the Society of Permanent Colors

Now that chemical colors deriving from processing petrochemical synthesis can be reproduced limitlessly, there has been a revolution in perception, which initially struck the eye of the viewer but which later accustomed the viewer's eyes and mind to their fixity and unchanging nature. The environment has again taken on stable yet inert colors.

The invention of the tin tube replacing an animal's bladder, together with chemical colors, has revolutionized painting, making it more convenient *plain air* (outdoors) and allowing the experimentation of the *pointillists*, in the same way as industrial enamel paints have made possible the *dripping style* painting of Pollock (1912–1956) [43], while quick-drying acrylic colors have given rise to a certain taste for Warhol's (1928–1987) [44] and pop art's wide-ranging, flat, uniform painting style.

It can be claimed that “abstract art was born with synthetic color, which became the most important and absolute principle of abstraction” [45].

While Max Bill (1908–1984) [46] sought chromatic harmony in laws of spatial values and in laws of color movement, distributed on the basis of proportions, Hartung (1904–1989) [47] was inspired by him in his appreciation of the energy of vinyl colors, which were scratched with different instruments and sprayed with compressors.

Later, after the invention of pearlescent and iridescent paints containing flakes of mica coated in titanium oxides, American minimalist painters covered surfaces with new changeable metallic

colors, freed from fixatives. Today, *acrylic* resin and silicone made possible the work of Pesce (1939) [48], where the colors blend in three dimensions.

In this concept of harmony, colors are being considered all the more beautiful when they are able to create a perceptive illusion and hide the nature of the material under a uniform permanent patina.

Color Harmony in the Society of Permutable Colors

Nowadays, digital screens are shifting vision from substance to light, prompting an adjustment toward a chromatic panorama characterized by intangibility and exchangeability.

In art, colored lights alter the perception of the physical place and introduce another space-time dimension. While Flavin (1933–1996) [49] was among the first to compose environmental works using fluorescent neon tubes, in which floors, walls, and ceilings seemed canceled by the colored light, embracing viewers in a suspended dimensions, nowadays digital media makes limitless experimentation possible. Viola (1951) [50], a veritable maestro, treats his works with a chromatic pattern making them similar to a painting, with a soft painting-like light running through them, a light that immerses the viewer in an emotional, sensorial experience taking place in an intermediate, part-natural and part-artificial space, which is both real and unreal. Toderi (1963) [51], an up-and-coming Italian video artist, transforms her urban film shooting into luminous stellar spaces, linking heaven and earth in a pulsating flow.

Color harmony nowadays is becoming dematerialized, exchanging artificial light which cancels materials, space, and time.

Conclusion

Giulio Carlo Argan wrote that “recognizing beauty is an act of justice and a sentimental act, meaning an act which confers value” [52].

In this way, as all that relates to matter of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – so also the definition of chromatic harmony is related to culturally accepted notions of taste and beauty. Given the changeable nature of notions such as beauty, taste, and aesthetic pleasure, any models to predict such are open to periodic renewal, creating new models or codes for new contexts.

Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Preference](#)
- ▶ [Color Scheme](#)
- ▶ [Colorant](#)
- ▶ [Colorant, Natural](#)
- ▶ [Inorganic, Hybrid, and Functional Pigments](#)
- ▶ [Maxwell, James Clerk](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Natural Color Distribution](#)
- ▶ [Neon Lamp](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Pearlescence](#)
- ▶ [Unique Hues](#)
- ▶ [Visual Illusions](#)

References

1. O'Connor, Z.: Color harmony revisited. *Color. Res. Appl.* **35**(4), 267–273 (2010)
2. Rumford, T.B.: *Recherche sur la Couleur*. Paul Migne, Paris (1804)
3. Goethe, J.W.: *Zur Farbenlehre*. Cotta, Tübingen (1810)
4. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colours*. Bell and Daldy, Cambridge (1879)
5. Noon, P., et al.: Delacroix. In: *Crossing the Channel: British and French Painting in the Age of Romanticism*, p. 58. Tate Publishing, London (2003)
6. “Seurat”: *Random House Webster’s Unabridged Dictionary*, New York (2007)
7. Ferretti-Bocquillon, M., et al.: *Signac, 1863–1935*. The Metropolitan Museum of Art, New York (2001)
8. Hölzel, A.: *Lehre von der harmonischen Äquivalenz*. Longen, München (1910)
9. Kandinsky, W.: *The Art of Spiritual Harmony*. Houghton Mifflin, London (1914)
10. Klee, P.: *Pedagogisches Skizzenbuch*. Longen, München (1925)
11. Itten, J.: *The Art of Color*. Van Nostrand Reinhold, New York (1961)
12. Albers, J.: *Interaction of Color*. Yale University, New Haven (1975)
13. Moholy-Nagy, L.: *Vision in Motion*. Theobald, Chicago (1961)
14. Droste, M.: *Bauhaus 1919–1933*. (Bauhaus Archiv) – Taschen/Vince K (2003)
15. Krawkow, G.V.: *Das Farbsehen*. Akad Verlag, Berlin (1955)
16. Newton, I.: *The principia: mathematical principles of natural philosophy*. University of California Press, Berkeley (1999)
17. Hoffmann, J.L.: *Versuch einer Geschichte der malerischen Harmonie überhaupt und der Farbenharmonie insbesondere*. Hendel, Halle (1786)
18. Runge, P.O.: *Die Farbenkugel*. Nossack, Hamburg (1810)
19. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
20. Plochere, G.: *Plochere Color System*. Vista, Los Angeles (1948)
21. Pfeiffer, H.E.: *L’Harmonie des Couleurs*. Dunod, Paris (1966)
22. Rosentsthiel, C.: *Traité de la Couleur*. Dunod, Paris (1934)
23. Fechner, G.T.: *Element der Psychophysik*. Breitkopf und Härtel, Leipzig (1860)
24. Munsell, A.H.: *Munsell Book of Color*. Munsell Colour, Baltimore (1942)
25. Munsell, A.H.: *A Grammar of Color*. Van Nostrand Reinhold, New York (1969)
26. Moon, P., Spencer, D.E.: Area in color harmony. *J. Opt. Soc. Am.* **34**, 93 (1944)
27. Moon, M., Spencer, D.E.: Aesthetic measure applied to color harmony. *J. Opt. Soc. Am.* **34**, 4 (1944)
28. Dimmick, F.L.: The dependence of auditory experience upon wave amplitude. *Am. J. Psychol.* **45**, 463 (1933)
29. Boring, E.G.: *History of Experimental Psychology*. Apleton-Century-Croft, New York (1950)
30. Mori, N., Nayatani, V., Tsujimoto, A., Ikeda, J., Namba, S.: An appraisal of two-colour harmony by paired comparison method. *Acta Chromatica* (Tokyo). **1**, 22 (1967)
31. Katz, D.: *The World of Colour*. Kegan Paul, Trench (1935)

32. Gelb, A.: Die "Farbenkonstanz" der Sehdinge. In: Bethe, A., v. Bergmann, G., Embden, G., Ellinger, A. (eds.) *Handbuch der Normalen und Pathologischen Physiologie*, pp. 594–678. Springer, Berlin (1929)
33. Granit, R.: *Sensory Mechanisms of the Retina*. Oxford University Press, London (1947)
34. Matiss, H.: *Les Notices d'une Peintre*. Dounod, Paris (1908)
35. "Piet Mondrian", Tate gallery, published in Ronald Alley, *Catalogue of the Tate Gallery's Collection of Modern Art other than Works by British Artists*, London (1981)
36. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York (1993)
37. Nemcsics, A.: Experimental determination of laws of color harmony. Part 1–8. *Color Res. Appl.* **32**(6), 477–488 (2007); **33**(4), 262–270 (2008); **34**(1), 33–44 (2009); **34**(3), 210–224 (2009); **36**(2), 127–139 (2011); **37**(5), 343–358 (2012), online 2013
38. Nemcsics, A.: Coloroid colour system. *Color. Res. Appl.* **5**, 113–120 (1980)
39. Hard, A., Sivik, L.: A theory of colors combination – a descriptive modes related to the NCS color order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
40. Brandi, C.: *Art Italian*. Art, Milano (1985)
41. Hume, D.: *Natural History of Religion*, Edinburgh (1952)
42. Alberti, L.B.: *De Pictura*, Basel (1540)
43. Polloch, P.J.: *Pollock-Krasner House & Study Center*, New York (2000)
44. Warhol, A.: *The Philosophy of Andy Warhol (From A to B & Back Again)*. Harcourt Brace Jovanovich, New York (1975)
45. Brusatin, M.: *Colore Senza Nome*. Marsilio, Venice (2006)
46. Bill, M.: *Funktion und Funktionalismus*. Schriften 1945–1988. Benteli, Bern (2008)
47. Hartung, H., Alley, R.: *Hans Hartung*. Oxford Art Online (2001)
48. Pesce: *Design Excellence Award of the Philadelphia Museum of Art* (2005)
49. Flavin, D.: *The Complete Lights, 1961–1996* by Michael Govan and Tiffany Bell, p. 49. Yale University Press, New Haven (2004)
50. Bill Viola: *The Eye of the Heart*. Dir. Mark Kidal [DVD]. Film for the Humanities & Sciences (2005)
51. Grazia Toderi. *Charta*, Milan (2004)
52. Argan, G.C.: *Design e mass media*. Op.cit. (Milan) (1965)
53. Luzzatto, L., Pompas, R.: *COLORI e MODA*. Bompiani, Milano (2018). ISBN 978-88-452-9748-9

Color Identification

- [Color Constancy and Categories](#)

Color Idiom

- [Chinese Color Language](#)

Color Image Statistics

- [Color Scene Statistics, Chromatic Scene Statistics](#)

Color Imaging

- [Hyperspectral Imaging](#)

Color in Care Environments

- [Color in Hospitals](#)

Color in Clinical and Nonclinical Areas

- [Color in Hospitals](#)

Color in Educational Environments

- [Color in Schools and Offices](#)

Color in Healthcare Environments

- [Color in Hospitals](#)

Color in Hospitals

Nilgün Olguntürk¹, Rengin Aslanoğlu¹ and Begüm Ulusoy²

¹Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

²Interior Architecture and Design, School of Design, University of Lincoln, Lincoln, Brayford Pool, UK

Synonyms

Color in care environments; Color in clinical and nonclinical areas; Color in healthcare environments

Definition

An aspect of hospital buildings in promoting the sense of well-being and comfort for all users besides the excellence of medical care

Introduction

Both the clinical and nonclinical areas of hospitals are created to promote the recovery process of patients by enhancing a sense of well-being for all users. The ambience of physical environments in hospitals – such as the entrance, reception areas, waiting areas, corridors, consulting rooms, and wards – can be attractive and provide visual stimulation. Creating ambience in such physical environments involves careful consideration of surface materials and paints, furniture, decoration, natural light, and artificial lighting. These diverse elements share an inseparable quality: color. Color affects user perception of the surroundings, thereby reducing general levels of stress especially by encouraging positive distraction and social support [1] as well as promoting confidence and safety. It serves in spatial orientation as a powerful tool used in coding, navigation, and

wayfinding [2]. Color also affects patients' recovery rates by improving the quality of experiences of patients, staff, and visitors [3]. Color design can be utilized for the benefit of patients, staff, and visitors in locations mutually occupied by persons of all ages who also have varying levels of health and mobility [2, 4].

Role of Color Application in Hospitals

An appropriate application of color can succeed in establishing a friendly, welcoming atmosphere comfortable for all [2]. Although hospital design necessarily should incorporate ease of access and inclusion of all kinds of persons, special attention needs to be paid to elderly as well as visually impaired persons, for whom ensuring optimal color application in their environment is crucial. Color applications are mainly important for the elderly and visually impaired people as such persons require clearer cues and need more time to comprehend visual information. Such users may also be less confident about occupying those environments [2, 5]. For both the elderly and the visually impaired, colors and textures of walls and floors are of the utmost importance in their navigation within the hospital environment without stress and enabling that they can develop a mental map for orientation in their surroundings [2]. Color contrast also needs to be provided in order to prevent accidents and establish spatial clues for wayfinding. Some of the colors, especially blue and green, become difficult to discriminate since human eye lens transmittance decreases with increased age [6]. Thus, color applications in hospitals should be carefully done for better color discrimination. For elderly and visually impaired people, color combinations with yellows, oranges, and reds can be used to improve visual performances.

Likewise, just as the elderly and visually impaired, children need to be able to understand their surroundings both intuitively and cognitively. Young children make associations with color more than forms. As Dalke et al. [2] indicated, children are less conscious about form but they can easily grasp colorful objects. This means that designing

environments with appropriate colors is especially important in children's hospitals. Such an environment should not only be interesting to children, but also created at their scale since children often experience fear and boredom in hospitals. Colorful and interesting playground areas can be positively distracting and serve to reduce negative emotions and lower stress levels.

Color can be used practically and functionally in accommodations and for wayfinding for patients and visitors, serving to make such spaces and orientation even more interesting. It is often difficult to navigate in such an unfamiliar environment. This difficulty can cause stress. As already mentioned earlier, color can be applied in signage, making wayfinding and mental mapping easier (Fig. 1). Especially, both the visitors and patients who enter the hospital may be understandably upset and preoccupied because of the urgency of their situation, and therefore, special attention needs to be given to establishing clear accessibility, signage, and other clues for orientation [2]. As well, generally, a hospital is complex, which makes wayfinding difficult [4]. Color coding can be employed to institute navigation routes, spatial hierarchies, and differentiation among facilities, a tool which enhances visual perception, visual attention, cognitive mapping, and the understanding of form. According to Dalke et al.

[2, 7], in coding colors, the variety of colors should be limited to minimize wayfinding confusion. A neutral color palette can be used since it directs user attention to accent colors in signage, landmarks, and important junction spaces [8]. However, study conducted with elderly people indicated that neutral color schemes (gray and beige) decrease the attention, whereas tones of blue and red enhance their navigation performances [9].

Reception, entrance, and waiting areas are important spaces where users can accommodate themselves to the hospital environment. Reception and entrance areas are perceived first, and when distinguished as landmarks, they can immediately help users to get oriented, e.g., through the usage of a strong color behind the receptionist [2] and distinctive colors in the entrance area. However, using dark colors are not advisable since they can negatively affect perception in narrow spaces and may generate confinement sensation [6]. An optimal choice of furniture, color, and lighting in entrance and waiting areas will not only serve to reduce stress [10], but also will create a comfortable, homelike atmosphere.

Care areas should be reassuring and provide a sense of excellent medical care combined with physical and emotional comfort [10]. Sensations of relaxation, comfort, cleanliness, and hygiene are important in hospital environments. In contrast

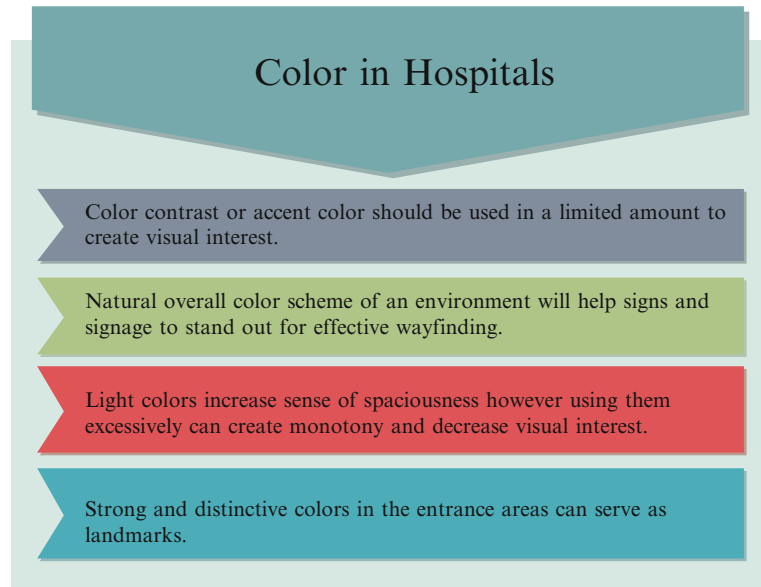
Color in Hospitals,

Fig. 1 İbn-i Sina Hospital, Ankara. Rengin Aslanoglu, 8 May 2018



Color in Hospitals,

Fig. 2 Summary of color applications in hospitals (figure created by the authors)



to common usage and a widespread preference for using white in the treatment of all surfaces of care areas, material and color variety can create a home-like atmosphere that promotes a sense of well-being. At the same time, using lighter colors is more necessary in narrow spaces [6]. Light colors increase sense of spaciousness which enhances well-being. As Delcampo-Carda, Torres-Barchino, and Serra-Lluch [6] highlighted, using achromatic colors on most of the surfaces can create monotony which negatively effects the recovery rates of patients and also can decrease morale of staff and visitors due to lack of stimulation and interest. Thus, colors can be used to give character to a space [6] to assist users' decision-making processes while solving a spatial problem related with wayfinding [9].

Color is an intrinsic characteristic of every material, and it is an inseparable tool in design. It literally covers and permeates all materials and its surfaces. Entailing everything from all materials, light, paint, art, and ambience, color is an essential component of the environment both aesthetically and functionally. It affects perception, emotions, and well-being. Color enhances the quality of life, i.e., variation of the hue or saturation of color has a positive influence on mood and human behavior. Color improves ambience and thereby increases the quality of experience, and it also reduces stress and creates safe

environments since color eases navigation, wayfinding, and the communication of information through color coding and the application of color in signage in public environments. Color, as one of the foremost features of hospital buildings (Fig. 2), has a role in patients' and other users' well-being and overall experiences in healthcare environments.

Cross-References

- [Achromatic Color](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Color Palette](#)
- [Color Scheme](#)
- [Environmental Color Design](#)

References

1. Andrade, C.C., Devlin, A.S., Pereira, C.R., Lima, M. L.: Do the hospital rooms make a difference for patients' stress? A multilevel analysis of the role of perceived control, positive distraction, and social support. *J. Environ. Psychol.* **53**, 63–72 (2017)
2. Dalke, H., Little, J., Niemann, E., Camgoz, N., Steadman, G., Hill, S., Stott, L.: Colour and lighting in hospital design. *Opt. Laser Technol.* **38**(4–6), 343–365 (2006)

3. Mahnke, F.H., Mahnke, R.H.: Color and Light in Man-Made Environments. Van Nostrand Reinhold, New York (1987)
4. Kumoğlu, Ö., Olguntürk, N.: The aid of colour on spatial navigation: a study in a virtual hospital environment. In: Proceedings of the European Conference on Cognitive Ergonomics, pp. 1–3 (2016)
5. Crews, J.E., Campbell, V.A.: Health conditions, activity limitations, and participation restrictions among older people with visual impairments. *J. Vis. Impair. Blind.* **95**(8), 453–467 (2001)
6. Delcampo-Carda, A., Torres-Barchino, A., Serra-Lluch, J.: Chromatic interior environments for the elderly: a literature review. *Color. Res. Appl.* **44**(3), 381–395 (2019)
7. Dalke, H., Camgöz, N., Bright, K., Cook, G., Littlefair, P., Loe, D.: Research with users: colour design and lighting for public transport, prison and healthcare environments. In: Proceedings of Include, Inclusive Design for Society and Business, Royal College of Art, London, CD-ROM, pp. 322–328. Ritsumeikan University, Kyoto (2003)
8. Malkin, J.: Hospital Interior Architecture: Creating Healing Environments for Special Patient Populations. Van Nostrand Reinhold, New York (1992)
9. Süzer, Ö.K., Olguntürk, N.: The aid of colour on visuospatial navigation of elderly people in a virtual polyclinic environment. *Color. Res. Appl.* **43**(6), 872–884 (2018)
10. Camgöz, N., Dalke, H.: Color design in healthcare environments. In: AIC Color 2005, Proceedings of the 10th Congress of the International Color Association, pp. 151–155. Comité Español del Color, Granada (2005)

Color in Schools and Offices

Nilgün Olguntürk¹, Rengin Aslanoğlu¹ and Begüm Ulusoy²

¹Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

²Interior Architecture and Design, School of Design University of Lincoln, Lincoln, UK

Synonyms

Color in educational environments; Color in workplaces

Definition

An aspect of school and office environments for improving visual processing, reducing stress, and enhancing task performances of all users.

Role of Color Application in Schools

Schools, where students spend most of their time, play an important role not only in their education but also in their physiological and psychological developments. Physical qualities of schools, described as the third educator, have vital effects on students' learning outcomes including place awareness [1]. With age, place awareness develops very rapidly [2, 3], as the young children are getting older, they start to obtain more information about their environments. Also, their abilities to understand and detect changes in the spaces improve too [4]. For an effective wayfinding, students should be able to relate spaces to other spaces and understand the basic layouts of space organizations, which often depend on the availability of environmental information [5]. Spatial learning begins in childhood but children need support while they are navigating themselves. One of the supports that can be provided by the physical environment is the usage of colors on surfaces, landmarks, furniture, etc. Besides its aesthetic and physical qualities, color is a powerful navigation tool.

The quality of school environment plays a significant role in students' way of learning [4, 6]. Hence, color in schools creates visual interest and it is one of the most important built environment elements in promoting navigation. Color in schools certainly improves children's wayfinding and spatial orientation abilities [7]. According to the findings of Read, children's sensory stimulation can be enhanced by using warm hues and bright accents in school environment surfaces [7]. It should be noted that color is a guiding tool, which should not dominate the environment [4]. Color can aid students' sense of security and safety since they need to know where they are in a complex building. Children's orientation and spatial cognition performances



Color in Schools and Offices, Fig. 1 Indoor landmark. Elif Helvacioğlu, 2007

can be improved by creating visual attraction points, i.e., indoor landmarks (Fig. 1) [4].

Color is not just the stationery of a surface [8]. By using color in classroom surfaces visual interest can be provided. Besides evoking visual interest, color can enhance children's cognitive performances. A study conducted by Brooker and Franklin [9] examined the effects of chromatic and achromatic colors in children's cognitive performances and found a positive impact of red in increasing their cognitive performances. Also it can be used to break monotony in classrooms where students easily get bored. Carefully applied and sparsely used colors can make classroom environments stimulating and supporting. In the classroom, side walls and the rear wall should have non-distinct colors where the front wall should benefit from a distinct color (Fig. 2). The purpose is to attract students' attention on what they have in their front and to rest their eyes when raising them up from the desk surface,

giving them a clear color contrast between the board, walls, materials, and the teacher [8].

In general, colors in a physical environment work as communication tools between people and the objects or surfaces around them [10]. It helps space recognition, encoding, and recalling while building mental maps of a physical environment. Color in schools (Fig. 3) provide optimum environments for children, which increases their visual processing, decreases stress, and improves brain development by creating challenges [11].

Role of Color Application in Offices

Colors are everyday interior architecture elements, which are inevitably affecting users. Kwallek et al. [12] highlighted their ubiquitous nature in our environments. Smith [13] discussed effects of colors on man-made environments such as interior architecture, architecture, urban design, etc., for their meanings and experiences in the context of COLPE (color, person, and environment) relationship and indicated that color, as an immanent environmental element, has an influential role. Kwallek et al. [14] underlined long hours that people spend in office spaces during their waking times, therefore their importance on users. Office spaces, in which people are spending their daytimes, have been in everyday life scene of people for many centuries. Moreover, office spaces are inherently evolving interiors depending on recent technologies and innovations that make them more ambiguous for color choices than other interiors. Contrary to other interiors, technologies and innovations can affect office spaces more directly and rapidly, since the function and duration of occupancy of technological developments are impactful in shaping the definition of office spaces. Last decade has witnessed not only massive usage of computers in office spaces but also different technologies such as mobile devices which provide more flexibility. Definition of office spaces are extended further and become part of different interior types in connection with many people starting to work from their homes or other public spaces. Fortunately, many enterprises do not miss the

Color in Schools and Offices, Fig. 2 Interior view of a classroom. Rengin Aslanoğlu, November 2013



Color in Schools

Color in schools creates visual interest, decreases stress, and improves brain development by creating challenges.

Carefully applied and sparsely used colors can make the classroom environments stimulating and supporting. In the classroom, side walls and the rear wall should have non-distinct colors where the front wall should have a distinct color.

Using warm hues and bright accents in school environment surfaces improves children's wayfinding and spatial orientation abilities.

Color in Schools and Offices, Fig. 3 Summary of color applications in schools. (Figure created by the authors)

opportunity and to satisfy their costumers' tendency to work out of their offices or homes, they include working/studying parts in their interiors.

Wall colors in interiors have been explored by different researchers in different contexts [15–18]. Similar to these studies, office interior colors

have been investigated by different researchers who focused on wall colors and their effects on occupants [12, 14, 19–22]. The most common and interesting feature of these studies is that they were conducted on a one-to-one scale experimental settings. The same feature was mentioned by Jalil et al. [23] that 11 of 12 studies on office/workspace, which the authors reviewed, had been conducted in one-to-one settings. These studies constructed the office environments on a one-to-one scale because it might be easier to simulate the offices' function and the activities done in those offices.

Previous studies demonstrated that colors have effects on arousal [14, 20, 21] mood [14, 24], productivity/performance [14, 20–22], appraisal [22], and satisfaction, and perceived performance [12] in office interiors. Kwallek et al. [19] showed a significant relationship between color of office interiors and number of errors during proofreading task in their study, which investigated nine different monochromatic colored offices. Their participants made more errors in office with white walls than blue one and red one separately, and light colors cause higher number of errors than darker colors in offices [19]. In addition, male participants felt more anger and depression in higher saturations whereas female participants felt more anger and depression in lower saturations [19]. Another previous study proved that environmental sensitivity and color schemes are an important determinant, which can affect mood and productivity of workers [14]. Low screeners, who cannot easily ignore irrelevant external elements in the environment, showed more dysphoria in red office than high screeners, who can eliminate irrelevant external elements [14]. White and blue-green office wall colors increase satisfaction and perceived job performance, and high and moderate screeners reported higher satisfaction and perceived performance [12]. Moderate screeners have enhanced performance than high and low screeners in blue-green office [20]. Küller et al. [21] compared three similar one-to-one scale experiments about office spaces: the first experiment investigated gray and colorful interiors, the second experiment explored red and blue interiors, and the last experiment investigated red and blue interiors for both clerical

and creative tasks. Their findings revealed that higher chromaticity increases perceived complexity and decreases perceived unity, blue interiors showed higher potency, and according to measurements, which were based on brain activities of participants during three experiments, colorful interiors are more arousing than red interiors which are more arousing than blue and gray interiors [21]. It is worth to indicate that personality (being introverted or extrovert) and nature of task (clerical tasks or creative tasks) are affected by interior colors differently and the authors proposed usage of warm colors to increase arousal. Öztürk et al. [22] proved that chroma has a significant effect on task performance and found that chromatic interiors cause more pleasant, more attractive, more satisfying, and more dynamic appraisals. Moreover, Öztürk et al. [22] stated that their findings about chroma effect on office interiors are parallel to findings of Küller et al. [21]: chroma has important effect on users' arousals in office interiors, achromatic office interiors are evaluated as more harmonious [22] and has higher unity [21], whereas chromatic ones are more dynamic [22] and complex [21]. Van der Voordt et al. [25] stated that the fundamental BOSTI study [26] revealed that low chroma colors were preferred more as office wall colors. According to Van der Voordt et al. [25] white is the most preferred color for both offices and meeting rooms, and male participants reveal higher percentage (43% and 32%, respectively) than female participants (38% and 28%, respectively), age and characteristic of person affect color preference in offices (Fig. 4). The study presented possible explanations of the preference of white and mentioned habituation might be one of them [25].

Previous studies have investigated chromatic, achromatic, and monochromatic color schemes in office interiors and stated that there are significant effects of color on office users' arousal, mood, performance, appraisal, and satisfaction and perceived task performance; however, these effects can differentiate according to characteristics of users, gender, and type of the task. One common discussion of these studies is the usage of white, which increases the number of errors [19] but are preferred more [25] and increases satisfaction and

Color in Schools and Offices, Fig. 4 Interior view of an office. Rengin Aslanoğlu, May 2018



Color in Offices

Color in office environments improve well-being, satisfaction, and performance.

White and blue-green office wall colors increase satisfaction and perceived job performance; workers in blue or red colored offices made less errors than white colored offices.

The effects of color schemes in office interiors can differentiate according to characteristics of users (their sensitivity towards environmental stimuli), gender, and type of the task.

Color in Schools and Offices, Fig. 5 Summary of color applications in offices. (Figure created by the authors)

perceived performance [12] in office interiors. Moreover, a recent study revealed that when white is used as a single wall color in an interior,

it cannot evoke any affective meaning, nonetheless, its pairs, separately with red and with green, can [18]. However, the previous study stated its

effect [19]; majority of users prefer white which should be investigated and clarified further because color can improve well-being, satisfaction, and performance in work environments [12, 19, 21, 27] and the preference of white might be a stereotype or a result of habituation for office interiors [25]. Van der Voordt et al. [25] mentioned color experience, mostly, unconsciously occurring in indoor environments, therefore, awareness of user might be low. Although, user preferences are important and valuable, interior architects should get over some learned and accustomed color usage in interior color schemes to overcome clichés, stereotypes, and preconceptions in order to provide more pleasant and more effective working environments in future office interiors, in working areas in public spaces, and in home-offices (Fig. 5).

Cross-References

- [Achromatic Color](#)
- [Color Contrast](#)
- [Color Scheme](#)
- [Environmental Color Design](#)

References

1. Berris, R., Miller, E.: How design of the physical environment impacts on early learning: educators' and parents' perspectives. *Australas. J. Early Childhood*. **36**(4), 102–110 (2011)
2. Hunt, E., Waller, D.: *Orientation and Wayfinding: A Review*. University of Washington Arlington, VA, (1999)
3. Matthews, M.H.: Environmental cognition of young children: images of journey to school and home area. *Trans. Inst. Br. Geogr.* **9**, 89–105 (1984)
4. Helvacioğlu, E., Olguntürk, N.: Colour contribution to children's wayfinding in school environments. *Opt. Laser Technol.* **43**(2), 410–419 (2011)
5. Helvacioğlu, E., Olguntürk, N.: Color and wayfinding. In: *Color and Light in Architecture First International Conference 2010, Proceedings* (2010)
6. Barrett, P., Zhang, Y., Moffat, J., Kobbacy, K.: A holistic, multi-level analysis identifying the impact of classroom design on pupils' learning. *Build. Environ.* **59**, 678–689 (2013)
7. Read, M.A.: Use of color in child care environments: application of color for wayfinding and space definition in Alabama child care environments. *Early Childhood Educ. J.* **30**(4), 233–239 (2003)
8. Niero, D., Premier, A.: Color in the schools. In: *Proceedings of the Color and Light in Architecture International Conference, Venice, Italy*, pp. 475–481 (2010)
9. Brooker, A., Franklin, A.: The effect of colour on children's cognitive performance. *Br. J. Educ. Psychol.* **86**(2), 241–255 (2016)
10. Hard, A., Sivik, L.: A theory of colors in combination – a descriptive model related to the NCS color-order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
11. Daggett, W.R., Cobble, J.E., Gertel, S.J.: Color in an Optimum Learning Environment, p. 1. International Center for Leadership in Education, Rexford (2008)
12. Kwallek, N., Soon, K., Woodson, H., Alexander, J.L.: Effect of color schemes and environmental sensitivity on job satisfaction and perceived performance. *Percept. Mot. Skills*. **101**(2), 473–486 (2005)
13. Smith, D.: Color-person-environment relationships. *Color. Res. Appl.* **33**(4), 312–319 (2008)
14. Kwallek, N., Woodson, H., Lewis, C.M., Sales, C.: Impact of three interior color schemes on worker mood and performance relative to individual environmental sensitivity. *Color. Res. Appl.* **22**(2), 121–132 (1997)
15. Wastiels, L., Schifferstein, H.N., Heylighen, A., Wouters, I.: Relating material experience to technical parameters: a case study on visual and tactile warmth perception of indoor wall materials. *Build. Environ.* **49**, 359–367 (2012)
16. Wastiels, L., Schifferstein, H.N., Heylighen, A., Wouters, I.: Red or rough, what makes materials warmer? *Mater. Des.* **42**, 441–449 (2012)
17. Ulusoy, B., Olguntürk, N.: Effects of color pairs on warmth perception in interiors. In: *Color and Imaging Conference, vol. 2016, No. 1*, pp. 126–135. Society for Imaging Science and Technology (2016)
18. Ulusoy, B., Olguntürk, N.: Understanding responses to materials and colors in interiors. *Color. Res. Appl.* **42**(2), 261–272 (2017)
19. Kwallek, N., Lewis, C.M., Lin-Hsiao, J.W.D., Woodson, H.: Effects of nine monochromatic office interior colors on clerical tasks and worker mood. *Color. Res. Appl.* **21**(6), 448–458 (1996)
20. Kwallek, N., Soon, K., Lewis, C.M.: Work week productivity, visual complexity, and individual environmental sensitivity in three offices of different color interiors. *Color. Res. Appl.* **32**(2), 130–143 (2007)
21. Küller, R., Mikellides, B., Jan, J.: Color, arousal, and performance – a comparison of three experiments. *Color. Res. Appl.* **34**(2), 141–152 (2009)
22. Öztürk, E., Yılmaz, S., Ural, S.: The effects of achromatic and chromatic color schemes on participants' task performance in and appraisals of an office environment. *Color. Res. Appl.* **37**(5), 359–366 (2012)

23. Jalil, N.A., Yunus, R.M., Said, N.S.: Environmental colour impact upon human behaviour: a review. *Proc. Soc. Behav. Sci.* **35**, 54–62 (2012)
24. Hsiao, L., Hsiao, M.C., Wang, Y.L.: Effects of office space and colour on knowledge sharing and work stress. *S. Afr. J. Econ. Manag. Sci.* **16**(5), 42–53 (2013)
25. Van der Voordt, T., Bakker, I., de Boon, J.: Color preferences for four different types of spaces. *Facilities*. **35**(3/4), 155–169 (2017)
26. Brill, M., Margulis, S.T., Konar, E.: *Using Office Design to Increase Productivity*, vol. 2. Workplace Design and Productivity, Buffalo (1985)
27. Lin, S.Y., Chang, H.I.: Does open-plan office environment support creativity? The mediating role of activated positive mood. *Creat. Stud.* **13**(1), 1–20 (2020)

Color in Transport

Nilgün Olguntürk¹, Rengin Aslanoğlu¹ and Begüm Ulusoy²

¹Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Çankaya, Ankara, Turkey

²Interior Architecture and Design, School of Design, University of Lincoln, Lincoln, Brayford Pool, UK

Synonyms

Color in transport hubs; Color in traveling

Definition

An aspect of transportation hubs to ensure wayfinding, comfort, safety, and orientation of all users.

Role of Color Application in Transport Hubs

Traveling is an exciting but mostly challenging experience which takes an individual from their current environment to another through different

spaces. Except traveling, people are moving through transport hubs for so many reasons such as work, receiving a medical treatment, education, etc., which might cause high level of stress and intolerance. From daily routines of undergrounds to life-changing journeys to other countries, transport hubs, which are a type of these in-between spaces, have crucial importance during departures and arrivals. Although, other transport hubs such as ferry quays are being used, previous research [1–3] mostly focused on ground transport and air transport because of their wide range of usage. Transport hubs should provide safety, security, comfort, clear access, well-defined data via signage or spatial elements, orientation information, and easy wayfinding features; therefore, their architectural and interior architectural elements require more attention in terms of design, lighting, material, and especially color, which is the inevitable part of proceeding elements. Previous studies have proved that colors have a fundamental influence on perception [4–7], meaning [8, 9], and understanding [10, 11] of different aspects in interiors. Occupants are exposed to color in transport hubs on big-scale surfaces which include different combination of color qualities under effects of diverse lighting conditions (Fig. 1). In transport hubs, where color coding is essentially necessary [2], colors are fundamental visual elements to be a source of information, to ensure wayfinding and orientation, and to contribute to glare control in the context of light and material relationship. Signage's color combination effects legibility and visual search performance which are the most crucial activities in transportation hubs [12].

Schindler [13] called for exploration of light-color-material interrelationship in contexts of time and place as a result of the interview with France and Michel Cler who were mentioned as architects of the 1980s and worked with large color plans and attached social approach to their works [14]. In the light of this study, Clers' view is salient and clear that color is a game changer for public spaces in different contexts: culture, material, industry, and urban spaces [13]. Colors appear in every interior and bring their both physical and metaphorical features with them. However, in an interior such as transportation hub

Color in Transport,

Fig. 1 Manchester Piccadilly Station. Begüm Ulusoy, 23 January 2018



which provides clues and data to their occupants, colors become the source of information and play key role on interior design contexts such as safety, user satisfaction, comfort, and perception. Dalke et al. [15] recommended that interior elements should provide information for wayfinding and orientation by color and contrast, and even staff need to wear uniforms which are easy to see. Colors arouse meanings in interiors; thus, some specific colors in specific environments can cause confusion, such as Dalke et al. [2] indicated that green signs can be misunderstood as “Fire Exit.” In addition, another study revealed that, in current conditions, retail signage is more visible and attention-grapping than signage of transportation hubs. The study recommends they should be matt in order to avoid glare which is harsh and uncomfortable bright light that is caused by the light source or its’ reflection interfering with visual perception [3, 16]. Color usage in interiors of transportation hub should be balanced in order to provide less glare with high visibility while providing accurate information.

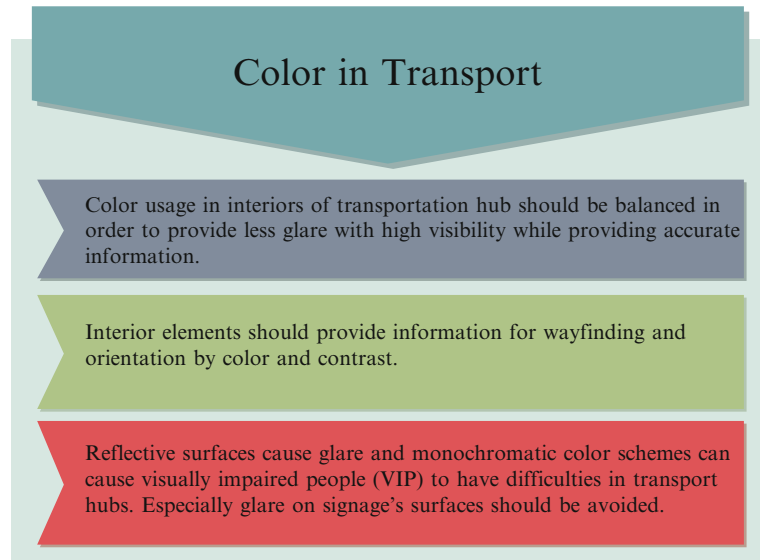
Cook et al. [17] investigated transport hubs such as train stations, bus stops, coach stops, underground, and airport; thus, they highlighted the vital importance of safety which can be maintained and supported by color, light, and design in interiors. In order to ensure comfort and safety, wayfinding and orientation have fundamental roles [2] which can be provided by color schemes in interiors. Wayfinding is a spatial problem-solving [18] which has been related to

colors [19]. Helvacioğlu and Olguntürk [19] demonstrated that colors are used as mental anchors in interiors to remember spatial features and to contribute to wayfinding. Reference points, which are mental anchors, reduce wayfinding difficulties in interiors [20]. Başkaya et al. [20] suggested colored waiting halls with other elements such as entrances, circulation systems as reference points for wayfinding, in the previous study for newcomers in an unfamiliar interior space. Colors can be used in different surfaces of interiors: walls, floors, ceilings, furniture, signage, etc. Previous study [2] pointed out light matt floor finishes in order to ensure contrast and highlighted the importance of signage’s shapes and colors rather than their texts for visually impaired people (VIP). As literature suggested, transport hubs should have a clear visual language, which should be designed and created by colors, in order to provide safety, comfort, and satisfaction for their occupants.

Previous study stated that major trends, such as wide usage of reflective surfaces or monochromatic color schemes cause visually impaired people (VIP) to have difficulties in transport hubs [3]. Glare as a major problem in such interiors for both VIPs and non-VIPs can be prevented by color and light design in interiors [1, 3]. Especially glare on signage’s surfaces should be avoided [3]; however, previous studies [2, 3] demonstrate that sparkle, i.e., attractive and extreme brightness on handrails may be useful for VIPs for their wayfinding and orientation [16]. Even though

Color in Transport,

Fig. 2 Summary of color applications in transport hubs. (Figure created by the authors)



they suggested that sparkle on handrails contributes to wayfinding and orientation of VIPs, Dalke et al. [2] stated that harsh glare on surfaces should be avoided; thus, extremely polished surfaces are not recommended. Colors which are inherently part of materials may contribute to glare control in interiors with lighting and material choices. They might be used to reduce glare in surfaces and to prevent any difficulties for the vision of occupants. In interiors, architects, interior architects, and designers aim to prevent glare; however, sparkle might be used, to the accompaniment of color, as an interior design element to clarify and highlight information which are required for wayfinding through vertical circulation elements in transport hubs' interiors.

Colors are not only fruitful and decorative but also functional interior design elements of architecture; thus, their existence and usage become more critically sensitive in public spaces such as transport hubs, shopping malls, hospitals, government offices, education buildings, etc. They are used as information sources and for wayfinding and orientation referenced by all occupants independently whether they are VIP or non-VIP. Glare should be prevented in specific surfaces such as signage, but sparkle might be used as contributing design elements on handrails for wayfinding in transport hubs' interiors. All colors provide

meanings in different contexts and in different scales of design; however, color becomes a source of information and reference point in transportation hubs' interiors, thus requires more attention and understanding (Fig. 2). In order to provide accurate wayfinding and orientation schemes and prevent glare, researchers, interior designers, and architects should focus on the relationship of colors in interiors with both other colors and indoor environment conditions, particularly lighting conditions and material properties.

Cross-References

- [Color Scheme](#)
- [Glare](#)
- [Sparkle](#)

References

1. Cook, G.K., Yohannes, I., Dalke, H., Camgöz, N.: Lighting multi-modal transport intersections for visually impaired travellers. In: *Licht für den Menschen = Lighting for humans = Lumière pour l'homme: 10th European Lighting Conference*, Berlin, Germany (2005)
2. Dalke, H., Cook, G.K., Bright, K.T., Camgöz, N., Yohannes, I., Niemann, E.: Future integrated transport environments: color design, lighting and visual

- impairment. Department for transport. In: Future Integrated Transport Programme: Progress and Results, pp. 32–33. Published Research Report. Department for Transport, London (2004)
3. Dalke, H., Camgöz, N., Cook, G.K., Bright, K., Niemann, E., Yohannes, I.: Future Integrated Transport Environments: Colour Design, Lighting and Visual Impairment. Department for Transport, London (2005)
 4. Ulusoy, B., Olguntürk, N.: Effects of color pairs on warmth perception in interiors. In: Color and Imaging Conference, vol. 2016, No. 1, pp. 126–135. Society for Imaging Science and Technology (2016)
 5. Odabaşıoğlu, S., Olguntürk, N.: Effects of coloured lighting on the perception of interior spaces. *Percept. Mot. Skills*. **120**(1), 183–201 (2015)
 6. Wastiels, L., Schifferstein, H.N., Heylighen, A., Wouters, I.: Relating material experience to technical parameters: a case study on visual and tactile warmth perception of indoor wall materials. *Build. Environ.* **49**, 359–367 (2012)
 7. Wastiels, L., Schifferstein, H.N., Heylighen, A., Wouters, I.: Red or rough, what makes materials warmer? *Mater. Des.* **42**, 441–449 (2012)
 8. Park, Y., Guerin, D.A.: Meaning and preference of interior color palettes among four cultures. *J. Inter. Des.* **28**(1), 27–39 (2002)
 9. Ulusoy, B., Nilgün, O.: Understanding responses to materials and colors in interiors. *Color. Res. Appl.* **42**(2), 261–272 (2017)
 10. van der Voordt, T., Bakker, I., de Boon, J.: Color preferences for four different types of spaces. *Facilities* **35**(3/4), 155–169 (2017)
 11. Kaya, N., Crosby, M.: Color associations with different building types: an experimental study on American college students. *Color. Res. Appl.* **31**(1), 67–71 (2006)
 12. Shi, Y., Zhang, Y., Wang, T., Li, C., Yuan, S.: The effects of ambient illumination, color combination, sign height, and observation angle on the legibility of wayfinding signs in metro stations. *Sustainability*. **12**(10), 4133 (2020)
 13. Schindler, V.M.: ‘Polychrome environments’ at the Centre Pompidou in Paris: the work of France and Michel Cler, architect–color consultants. In: Proceedings of the International Interdisciplinary Conference on Color and Pattern Harmony, 11–13 June 2012, Óbuda University, Budapest, Hungary (2012)
 14. Serra, J., García Codoñer, A.: Color composition in postmodern western architecture. *Color. Res. Appl.* **39**(4), 399–412 (2014)
 15. Dalke, H., Camgöz, N., Cook, G.K., Yohannes, I., Bright, K.T.: Inclusive transport environments: color design and visual impairment. In: Proceedings of the 11th International Mobility Conference (IMC 11), Stellenbosch, South Africa (2005)
 16. Egan, M.D., Olgyay, V.: Architectural Lighting. McGraw-Hill, New York (2002)
 17. Cook, G.K., Dalke, H., Bright, K.T., Yohannes, I., Camgöz, N.: Inclusive transport environments for people who are visually impaired. In: Proceedings of the 11th International Mobility Conference, Stellenbosch, South Africa (2003)
 18. Passini, R.: Spatial representations, a wayfinding perspective. *J. Environ. Psychol.* **4**(2), 153–164 (1984)
 19. Helvacıoğlu, E., Olguntürk, N.: Color and wayfinding. In: Color and Light in Architecture First International Conference 2010 Proceedings (2010)
 20. Başkaya, A., Wilson, C., Özcan, Y.Z.: Wayfinding in an unfamiliar environment: different spatial settings of two polyclinics. *Environ. Behav.* **36**(6), 839–867 (2004)

Color in Transport Hubs

► Color in Transport

Color in Traveling

► Color in Transport

Color in Workplaces

► Color in Schools and Offices

Color Information Processing in Higher Brain Areas

Hidehiko Komatsu
Brain Science Institute, Tamagawa University,
Machida, Japan

Synonyms

Color processing in the inferior temporal cortex;
Color processing in the ventral higher visual area;
Cortical color processing

Definition

Ventral higher visual areas in the cerebral cortex play critical roles in color vision. Neural processing of color vision is studied in the inferior temporal cortex of the macaque monkey where multiple sites related to color vision are identified.

Human Studies Showing the Importance of the Ventral Higher Area in Color Vision

Damage to the human cerebral cortex can cause the loss of color vision (cerebral achromatopsia) [1]. Overlap of the damaged areas among many achromatopsia patients is the highest in a relatively small area in the ventral occipitotemporal cortex surrounding the fusiform and lingual gyri (Fig. 1). This area is located adjacent to the area where damage causes deficits in face perception (prosopagnosia). In functional-imaging experiments in human subjects with normal color vision, high activity during color perception is observed around this area [2, 3]. In addition, human subjects exposed to electrical stimulation to the color responsive sites in the ventral occipitotemporal cortex report experience of illusory color perception [4]. Together, these

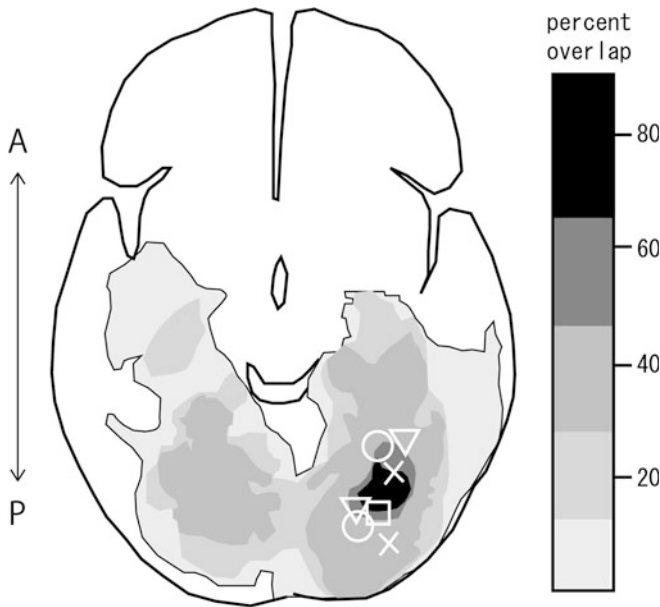
findings provide converging evidence that the ventral occipitotemporal cortex is critically involved in color perception. In functional-imaging studies, multiple peaks of activity to color stimuli are observed along the anterior-posterior direction in this area [3, 5], suggesting the existence of multiple subregions in this area that are related to color vision.

Importance of the Inferior Temporal Cortex in Color Vision in the Macaque Monkey

Neural processing of color information in higher visual areas was studied in macaque monkeys. Visual signals emanating from the primary visual cortex (V1) are processed along two cortical pathways: the ventral visual pathway and the dorsal visual pathway. Along the ventral visual pathway, color information processing starts in V1, is relayed by areas V2 and V4, and finally reaches the inferior temporal (IT) cortex [6]. The IT cortex can be divided into multiple subareas. One way is to assign the anterior two-thirds of the IT cortex as area TE, and the third as area TEO. TEO has a crude retinotopic organization, while TE does not. TE neurons tend to have large receptive fields

Color Information Processing in Higher Brain Areas,

Fig. 1 Overlap of brain damages among achromatopsia patients [1]. The scale indicates the degree of overlap. Symbols indicate locations of peak activations evoked by color perception in several functional-imaging experiments. Anterior is to the top, and posterior is to the bottom. (Adapted and modified from Bouvier and Engel [1])



including the fovea. TE can be further separated into an anterior and posterior half that are named anterior inferior temporal (AIT) cortex and central inferior temporal (CIT) cortex, respectively. In this case, TEO is named posterior inferior temporal cortex (PIT).

A widely held misconception is that area V4 is the center for color vision. A large bilateral lesion of V4 causes only mild impairments in color discrimination. It is more reasonable to consider V4 an intermediate stage of the ventral visual pathway that processes information related to object vision such as color, shape, and texture, and which transmits this information to the IT cortex.

Bilateral lesions of the IT cortex cause permanent deficits in color discrimination [6, 7]. Those monkeys exhibit significant elevation of the color discrimination threshold but less severe deficits in discrimination of gray in different luminance levels were observed. Because cortical damage in lesion studies encompasses a large area within the IT cortex, it is not possible to pinpoint whether a specific subregion is most responsible for the observed impairments. Nonetheless, attempts of partial ablation of the IT cortex showed that damage in the middle temporal gyrus is responsible for the color discrimination deficits. This is consistent with the localization of color selective activities at multiple sites in this region [6]. One study also showed that a lesion restricted to area TEO does not cause a deficit in color discrimination suggesting that area TE is most important for color vision in the macaque monkey.

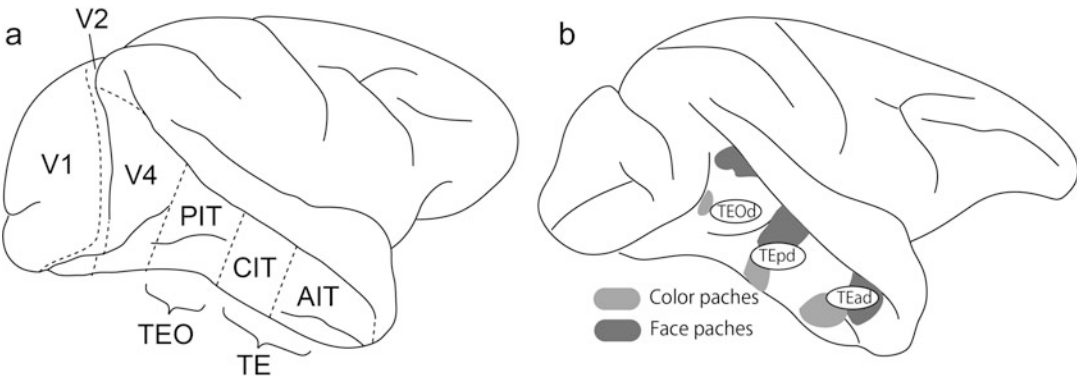
Functional Organization of the IT Cortex in Relation to Color Processing

There is some functional organization related to color processing in area V4 and the IT cortex. Functional magnetic resonance imaging (fMRI) studies have shown that several patchy regions of a few millimeter scale exist in V4 and in IT that are activated more strongly by chromatic stimuli than achromatic stimuli [5, 6, 8]. Electrophysiological recordings from some of these color preferential domains have shown that color-selective neurons are clustered in these regions

[8]. In the IT cortex, such color preferential domains are localized in the PIT, CIT, and AIT along the middle temporal gyrus [6]. Multiple face-selective regions exist in the IT cortex of the monkey where face-selective neurons are clustered. Color preferential domains are located adjacent and ventral to these face-selective domains [5, 6] (Fig. 2). A neuroanatomical study and an experiment combining fMRI and electrical stimulation have shown that at least some of these color preferential domains are mutually connected [9, 10]. These results suggest that color information is processed at least partially by neural pathways specialized for color information. In V4 and PIT, responses within the color-biased regions are less sensitive to stimulus orientation than those outside the color-biased regions. However, this does not imply that neurons in the color preferential domain carry only color information and do not encode information of other visual features such as shape and texture. Many color-selective IT neurons are also sensitive to shape information but there seems to exist a gradation in the separation of color and shape information in the IT cortex: neurons in the AIT cortex tend to exhibit more shape-invariant color selectivity than those in more posterior parts of the IT cortex [8, 10].

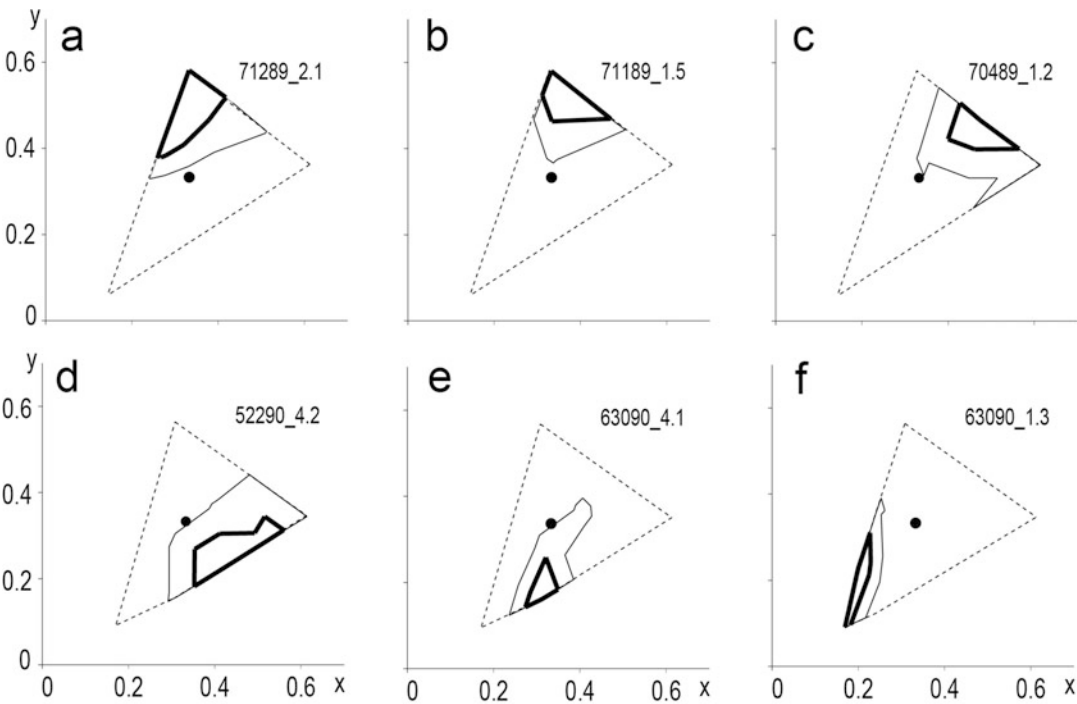
Neural Representation of Color in the IT Cortex

Color selectivity of IT neurons is analyzed by recording single-neuron activities and testing their responses to color stimuli based on the chromaticity diagram. Many color-selective neurons in both AIT and PIT are selective for particular hues and some are selective for saturation as well [8, 11, 12] (Fig. 3). There seems to exist a micro-organization of hue representation in the IT cortex: neurons preferring similar hues are arranged in a columnar fashion perpendicular to the cortical surface and the preferred hue gradually changes along the cortical surface [13]. Each color domain in the IT cortex contains neurons preferring various hues and population activities of neurons in each color domain systematically represent the entire hue circle [12]. Color selectivity of V4 and PIT neurons



Color Information Processing in Higher Brain Areas, Fig. 2 (a) Lateral view of the cerebral cortex of the macaque monkey. Visual areas along the ventral visual pathway are indicated. (b) Activation patterns to faces (dark gray) and colors (light gray) measured by fMRI in the macaque monkey [6]. In each panel, anterior is to the

right, posterior is to the left. IT, inferior temporal cortex; AIT, anterior inferior temporal cortex; CIT, central inferior temporal cortex; PIT, posterior inferior temporal cortex; TEad, dorsal subregion of the anterior TE; TEpd, dorsal subregion of the posterior TE; TEOd, dorsal part of area TEO. (b is adapted and modified from Conway [6])



Color Information Processing in Higher Brain Areas, Fig. 3 Color selectivity of six example neurons recorded from the AIT cortex tuned to different hues. Color selectivity of each neuron is represented as response contours within the region in the examined CIE-xy chromaticity

diagram (broken line). In each panel, the thicker (thinner) response contour indicates the locations where the responses were 75% (50%) of the maximum response. A dot at the center of each panel indicates a white point; adapted from Komatsu et al. [11] with modification

tend to be affected by the luminance contrast of the color stimuli, while those of AIT neurons are more invariant to the luminance contrast [12]. It appears that separation of luminance and color occurs in the AIT cortex, resulting in color representation invariant of the luminance contrast. It was also shown that color-selective neurons in the AIT cortex exhibit a significant correlation between the trial-to-trial fluctuations in neural responses and behavioral color judgments [8].

The AIT cortex has strong mutual connections with the prefrontal cortex that plays an important role in executive control of behavior. It was shown that color-selective responses of AIT neurons are affected by the ongoing task-demand such that the amplitude of the responses to color stimuli changes depending on the task demand (i.e., fine discrimination or coarse categorization of color stimuli) [8].

Unresolved Issue: Processing of Object Color

In natural and artificial environments, color vision is tightly related to the recognition of object identity or surface quality of an object. However, most studies on the role of the IT cortex in color vision have used color stimuli isolated from the visual context. Thus, our knowledge remains short of understanding the representation of object color. To address this problem, we need to know how the effects of illumination are discounted and how the optical properties of objects such as reflectance and transmittance are represented. These questions are related to a wider range of challenges such as color constancy, material perception, and intrinsic image decomposition. Some insights on these problems are available from previous studies. For example, double opponent neurons in V1 and silent suppressive surroundings of color-selective neurons in the visual cortex may be implicated in color constancy [8, 14]. In addition, it was shown that information related to material perception is mainly processed in the ventral visual pathway and that some neurons in the IT cortex are selective for surface reflectance [15]. However, more studies are needed to achieve

a comprehensive understanding of the neural processing of object color.

Cross-References

- [Color Processing, Cortical](#)
- [Magno-, Parvo-, and Koniocellular Pathways](#)

References

1. Bouvier, S.E., Engel, S.A.: Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cereb Cortex*. **16**, 183–191 (2006)
2. McKeefry, D.J., Zeki, S.: The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*. **120**, 2229–2242 (1997)
3. Bartels, A., Zeki, S.: The architecture of the colour centre in the human visual brain: new results and a review. *Eur J Neurosci*. **12**, 172–193 (2000)
4. Schalk, G., et al.: Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain. *Proc Natl Acad Sci U S A*. **114**, 12285–12290 (2017)
5. Lafer-Sousa, R., Conway, B.R., Kanwisher, N.G.: Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *J Neurosci*. **36**, 1682–1697 (2016)
6. Conway, B.R.: The organization and operation of inferior temporal cortex. *Annu Rev Vis Sci*. **4**, 381–402 (2018)
7. Komatsu, H.: Mechanisms of central color vision. *Curr Opin Neurobiol*. **8**, 503–508 (1998)
8. Conway, B.R., et al.: Advances in color science: from retina to behavior. *J Neurosci*. **30**, 14955–14963 (2010)
9. Banno, T., et al.: Reciprocal connectivity of identified color-processing modules in the monkey inferior temporal cortex. *Cereb Cortex*. **21**, 1295–1310 (2011)
10. Chang, L., Bao, P., Tsao, D.Y.: The representation of colored objects in macaque color patches. *Nat Commun*. **8**(1), 2064 (2017)
11. Komatsu, H., et al.: Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *J Neurosci*. **12**, 408–424 (1992)
12. Namima, T., et al.: Effects of luminance contrast on the color selectivity of neurons in the macaque area V4 and inferior temporal cortex. *J Neurosci*. **34**, 14934–14947 (2014)
13. Conway, B.R., Tsao, D.Y.: Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proc Natl Acad Sci U S A*. **106**, 18034–18039 (2009)

14. Komatsu, H.: The physiological substrates of color constancy. In: Walsh, V., Kulikowski, J. (eds.) *Perceptual constancy*, pp. 352–372. Cambridge University Press, Cambridge (1998)
15. Komatsu, H., Goda, N.: Neural mechanisms of material perception: quest on Shitsukan. *Neuroscience*. **392**, 329–347 (2018)

Color Language

- [Crowdsourcing of Multilingual Color Names](#)

Color Language Translation

- [Translation of Color Language](#)

Color Lexicon

- [Chinese Color Language](#)

Color Magnitude Diagrams

Michael H. Brill
Datacolor, Lawrenceville, NJ, USA

Definition

A color-magnitude diagram is a scattergraph of astronomical objects showing the relationship between each object's absolute magnitude and its estimated surface temperature or between optical or perceptual proxies for these quantities.

Historical Antecedents

Humankind has always wanted to understand the bodies in the night sky, and one step to understanding them is to categorize them.

The first tool available to assign these categories was the human visual system – the unaided eye. Hipparchus (c. 190 BCE–c. 120 BCE) developed a scale for stars based on visual brightness, which eventually became quantified as stellar magnitude. The convention that dimmer stars have higher magnitude is a historical precedent that dates from Hipparchus [1]. However, whereas Hipparchus attached magnitude 1 to the brightest star within each constellation, Ptolemy (c. 140 AD) refined the system so that the brightest stars had magnitude 1 and the barely visible stars had magnitude 6 [1].

From the time of Galilei (1564–1642) and Kepler (1571–1630), the unaided eye became augmented by telescopes, whose main function in astronomy is to gather light from a much larger area than is available at the pupil of the eye. Telescopes were at first very limited in their light-gathering ability but later developed enough aperture so that the stars became accessible to color vision. At that point, a new categorization was possible, in which color and brightness could be conjoined.

Of course, stars have spectra that are close to black-body radiators, and such radiators have only two physical variables: intensity and temperature [2]. In perceptual terms, the color ranges from reddish through bluish through yellow and white, on a curve in the chromaticity diagram called the black-body locus. It is therefore possible to render the visible character of a star by two numbers: the brightness (magnitude) and a single variable of color (or temperature).

Diagram for Stars

At the beginning of the twentieth century, aided by the mature technology of telescopes, Danish astronomer [3] and American astronomer [4] developed the first color-magnitude diagram, called the **Hertzsprung–Russell diagram (H-R diagram)** [5–7]. Originally the diagram was based on visual estimation of magnitude and color, and it was a research tool to help characterize stellar evolution before the mechanism of nuclear fusion was understood. After about the 1930s, the H-R diagram became based on

objective measurements but was used less as a research tool and more as a way to illustrate the theoretically predicted evolution of stars through trajectories in the diagram.

The objective measurements that replaced direct human perception were as follows: A star's absolute magnitude is the attenuation (in factors $10^{-0.4}$) of the star Vega's power (as received at 10 pc viewing distance [32.6 light years]) to equal that of the star (also corrected to 10 pc).

A star's surface temperature is estimated in one of three ways: by the observed color (an old way), by a comparison of two sensor outputs such as blue and violet (a newer way), or by a model prediction of the temperature of a black-body radiator with the same radiation power per unit star-surface area (the most modern way). The third way requires independent inference of the star's radius but assumes the star has an emissivity of 1. To acknowledge the lack of compensation for the true emissivity, the temperature on an H-R diagram is called "effective temperature."

It is important to note the coordinate conventions of the H-R diagram: Temperature on a log scale [7] increases from right to left and magnitude (i.e., dimness) increases from bottom to top. Hence, dim red stars (red giants) appear near the upper right of the diagram, and bluish bright stars (such as white dwarfs) appear in the lower left. A long cluster of stars called the Main Sequence extends from the upper left to lower right. Higher-mass stars occur at the upper left of this sequence, and the Sun appears approximately in the middle. The Main Sequence is composed of stars that are currently dominated by hydrogen that is fusing into helium. According to currently accepted theory of stellar evolution, such stars will eventually migrate either to the red-giant or white-dwarf domain of the H-R diagram.

H-R diagrams are often depicted in color, either pseudo-color with a thermal code to show the temperature or coded according to star categories such as cluster membership. No matter how the measurements were obtained, the data representation for the end user returns to visual perception as the way to see the color and brightness relationships of stars.

Diagram for Galaxies

One further stage in telescope evolution occurred in 1990 when the Hubble Space Telescope was launched into orbit. The Hubble Space Telescope was able to render high-contrast images outside the Earth's atmosphere without encountering the absorption, scattering, and haze that beset Earth-bound telescopes. In digital images conveyed to earth from the Space Telescope, it was possible to see the colors, not only of stars, but of whole galaxies that are so distant as to be too dim to measure on the Earth. The color-magnitude diagram developed by Hertzsprung and Russell then evolved to accommodate galaxies. Thus was born the **galaxy color-magnitude diagram** [8]. The construction of such a diagram is similar to that of the Hertzsprung–Russell diagram, but the interpretation in terms of physical properties is not as precise.

To compute a galaxy's absolute magnitude, it is treated as a point-like object, whereupon its radiation is corrected (by the inverse-square law) to a distance of 10 pc, and the absolute magnitude is computed as the number of $10^{-0.4}$ attenuations of similarly compensated power from a reference to achieve a match.

The temperature is also the same "effective temperature" as is used in the H-R diagram. However, galaxies are visible from much farther away than individual stars, so a galaxy's recessional red shift exerts an appreciable influence on its effective temperature. Because light from farther galaxies (and greater red shift) requires more time to travel to the Earth, the galaxy color-magnitude diagram, if it were not red-shift-corrected, would depict an earlier universe at the red end of the temperature scale. However, in practice, the temperature is red-shift-corrected to be in the "rest frame" of the galaxy. In this case, the galaxies divide into two clusters: bright and reddish as opposed to dim and bluish. The existence of such clusters places strong constraints on theories of galactic formation.

Color-magnitude diagrams are not always defined in terms of fundamental physical properties such as those described above. In one embodiment, the galaxy color-magnitude diagram coordinates are defined by the log outputs of two

sensors, typically called g and r , the sensors having different spectral sensitivities: the abscissa records r (which stands in lieu of the absolute magnitude) and the ordinate records $g - r$ (which relates to the temperature). At the top of the diagram, one finds the red sequence of galaxies (elliptical galaxies), and at the bottom, one finds the blue cloud (spiral galaxies).

Cross-References

- [Galaxy Color Magnitude Diagram](#)
- [Hertzsprung-Russell Diagram](#)

References

1. Astronomical Society of South Australia: Stellar Photometry. <http://www.assa.org.au/articles/stellarphotometry>. Accessed 30 May 2013
2. Wyszecski, G., Stiles, W.S.: Color Science, 2nd edn. Wiley, New York (1982)
3. Hertzsprung, E.: Über die Sterne der Unterabteilung c und ac nach der Spektralklassifikation von Antonia C. Maury. *Astron. Nachr.* **179**(4296), 373–380 (1909)
4. Russell, H.N.: Relations between the spectra and other characteristics of the stars. *Pop. Astron.* **22**, 275–294 (1914)
5. Gribbin, J.: Companion to the Cosmos. Little, Brown, New York (1996)
6. Pasachoff, J.M.: A Brief View of Astronomy. Saunders College Publishing, New York (1986)
7. Wikipedia, Hertzsprung-Russell Diagram: http://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram. Accessed 30 May 2013
8. Shapley, A.E.: Physical Properties of Galaxies from $z = 2-4$. <http://arxiv.org/abs/1107.5060v2>. Accessed 26 Feb 2013

Color Mapping

- [Color Transfer](#)

Color Mixture

- [Color Combination](#)

Color Models

- [Color Order Systems](#)

Color Name Applications in Computer Vision

C. Alejandro Parraga¹ and Arash Akbarinia²

¹Computer Vision Centre/Computer Science Department, Universitat Autònoma de Barcelona, Bellaterra, Spain

²FB06 Psychologie und Sportwissenschaften, Justus-Liebig Universität Giessen, Giessen, Germany

Synonyms

[Chromatic description](#); [Chromatic labeling](#); [Color categorization](#); [Color category boundaries](#); [Color names for image understanding](#); [Color terms](#); [Semantic-color interaction](#)

Definition

Humans exchange color information through the use of language, associating special words (color names) to categorical regions of color space. Color name applications in computer vision refers to algorithms that process digital images, labeling each pixel or pixel region with a given word or combination of words that matches the color name a human would give to such pixel or region.

The Problem of Labeling Image Regions by Color Name

Items belonging to different categories are easier to discriminate from each other than items belonging to the same category along a given perceptual dimension. In the case of color, the human visual system enhances the separation between certain

regions of the color space so that some colors are perceived sharply different from its neighbors, as is the case of the “bands” we see in the rainbow. These regions are then semantically categorized and assigned color names to aid communication. There are many computational applications where we need to accurately replicate these semantic categorizations to have a performance similar to humans.

The task of associating names to colors seems extremely easy for us humans, until we are presented with (semantic) categorical boundaries. This is an all too familiar situation: We ask for a yellow paint, and the shopkeeper brings us a brown-looking sample; we strongly disagree with friends about the color of our new car; or we are dumbfounded by the amount of different names given to “pink” in the paint samples’ catalogue. Providing that viewing conditions are the same, it is safe to assume that healthy individuals perceive the same colors, (This is not strictly true, and there are some astonishing but very limited exceptions such as #thredress [1].) but still, there are regions of the color space where individual observers often assign different names to the same color. In summary, there is strong agreement when naming some regions of the color space, and strong disagreement when naming others. The former are often called “focal colors,” selected by observers to best represent a given color name (e.g., “red” or “purple,” etc.), and the latter are often referred to as “categorical boundary” colors. The precise source of these individual differences remains unknown, although some authors point out the differences in the physiology (macular and lens density, pigmentation density, etc.), environmental factors (geographical differences in the environment’s statistics), linguistic and cultural factors, etc. Perhaps the most popular and influential study of basic color names across cultures and languages is that of Berlin and Kay [2], who proposed the existence of 11 “basic color terms” (in English they are “black,” “blue,” “brown,” “green,” “gray,” “orange,” “pink,” “purple,” “red,” “white,” and “yellow”). This study was later supported by others [3]. Although this and most of the research involve focal colors, some researchers have explored and mapped the

boundaries between categories trying to relate their findings to the subjacent neurophysiology [4, 5, 6]. Other factors such as viewing conditions, illumination, or the influence of surrounding areas can modify not only the perception of a given color but also category assigned.

Why Label Image Regions by Color Name?

There are many tasks we want computers to perform better than humans (e.g., driving; recognizing objects, gait, or faces; tracking; translating languages; etc.), and there are others where human performance is the milestone. In other words, we want algorithms to include the variability which is intrinsic in humans (e.g., art appreciation, image understanding, content categorization, etc.). In many such tasks, extracting higher level descriptors that provide clues to image content such as the names of the colors is fundamental. These descriptors can be later combined with image segmentation, used to select objects by color, describe the appearance of the image, generate semantic annotations, etc. A typical example is image retrieval, where the user enters a semantic description of the image (objects content, size, position, main colors, etc.), and the algorithm retrieves the most likely candidates from a large image database. Nowadays, when human-machine interaction is ubiquitous across a wide range of applications, its success ultimately requires a smooth integration with color naming: We do not want to miss possible candidates because the algorithm is constrained to a tight definition of color categories. The same applies when semantically describing the content of the picture or classifying it automatically (e.g., images of brown cars versus images of beige cars). Researchers have made use of many techniques to solve these problems.

Computational Solutions

Broadly speaking, there are two kinds of computational solutions: those based on a psychophysically based color-space partition and those

developed by applying machine learning to natural images.

Space Partitioning-Based Models

The easiest computational approach to semantically labeling colors is to create “dictionaries” of exemplary colors and names [7] or to partition the color space in regions corresponding to an arbitrary number of color names. This fairly rigid approach works reasonably well for high agreement regions (focal colors) but is not good enough to capture the large variability of human percepts.

In 1978, Kay and McDaniel went further, proposing a model based in Fuzzy set theory (which in turn is a development from standard set theory) [4] that considers four fuzzy sets (“red,” “green,” “yellow,” and “blue”). The degree to which each pixel was a member of a particular semantic category was specified as a value between zero and unity. In other words, pixels were labeled with a numerical vector corresponding to their “belonging” to these predefined semantic categories. In this framework, universal “focal” color regions can be understood as regions of maximum color-category membership functions (i.e., unity). Correspondingly, nonfocal colors have positive but nonmaximal degrees of semantic category membership and are members of more than one basic category. For the English language, these include colors such as yellow-green, reddish-purple, etc. These nonfocal, multiple membership colors capture the variability in boundary judgments reported by Berlin and Kay [2]. The membership functions of the semantic color categories were derived from physiological neural response functions for wavelength and fuzzy set operations (union and intersection). Using this model, they reinterpreted the evolutionary sequence of basic color names proposed by Berlin and Kay as the successive refinements of previously existing basic color names. A further extension was proposed by Lammens [8] who fitted Gaussian normal distributions to the 11 basic colors. This same paradigm was developed a step further by Benavente et al. [9] who incorporated arbitrary “fuzzy” membership regions (combinations of Sigmoids and Gaussians) to fit B&K’s categories to Surges and

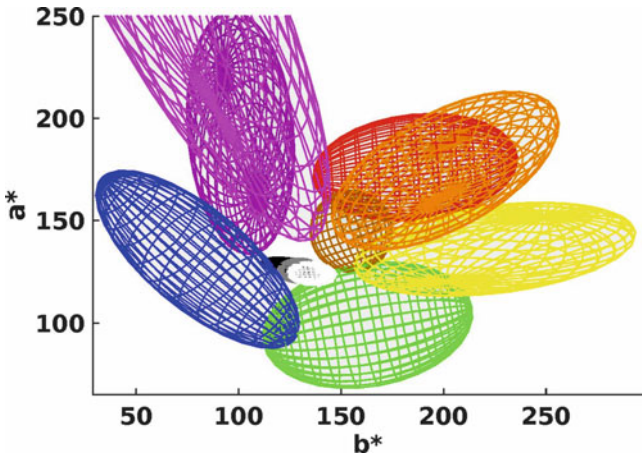
Whitfield’s psychophysical data [10] over three lightness levels in CIELab. Mojsilovic [11] proposed a geometrical solution in CIELab that assumed a well-represented set of prototypes (foci), computing the distance between a given color and all prototypes to obtain membership values. Seaborn et al. [12] developed a model based on a Fuzzy C-Means clustering algorithm applied to psychophysical measurements by finding local minima within the group sum of squared error objective function. Menegaz et al. [13] proposed a three-dimensional Delaunay triangulation of the color space to fit their own color name data to vertices of 3D tetrahedra in CIELab space. The final 11 fuzzy categories were obtained by a series of geometrical interpolations. All the previous models were adjusted to regions of high color-category membership (maximum agreement among observers), using linear interpolation to obtain membership values outside these. A completely different approach was adopted by Parraga and Akbarinia [6] who hypothesized that linear interpolation might not be the best way to obtain categorical boundaries in CIELab and decided to psychophysically measure the performance of observers in the transitional regions. They quantified the variability and fitted the surfaces of 11 three-dimensional Gaussians (3D ellipsoids, see Fig. 1) to these boundaries. Table 1 provides a comparison of these color categorization models.

Random Initialization-Based Models

An alternative computational approach is by means of learning color names (categories) from “real-world” images. Following this line, Yendrikhovskij [17] reasoned that natural image’s hues belonging to the same color name should cluster together in a perceptually uniform space. Consequently, he applied k-means to acquire color categories from samples of pixels of natural images, using a minimum-distance criterion among members of the same color category. In a later work, probabilistic latent semantic analysis (PLSA) was applied to learn color names from a collection of images obtained from search engines [16]. PLSA was originally introduced for document analysis, and van de Weijer et al. [16]

Color Name Applications in Computer Vision,

Fig. 1 Schematics of the three-dimensional ellipsoids used by Parraga and Akbarinia [6] to partition CIELab color space into 11 categorical regions



Color Name Applications in Computer Vision,

Table 1 Comparative results of several color categorization models when applied to the Berlin and Kay [2] and Sturges and Whitfield [10] psychophysical results. Keys: LGM: Lammens’s Gaussian model [8]; MES: MacLaury’s English Speaker [14]; TSM: Benavente et al.’s Triple Sigmoid model [9]; SFKM: Seaborn’s fuzzy k-means model

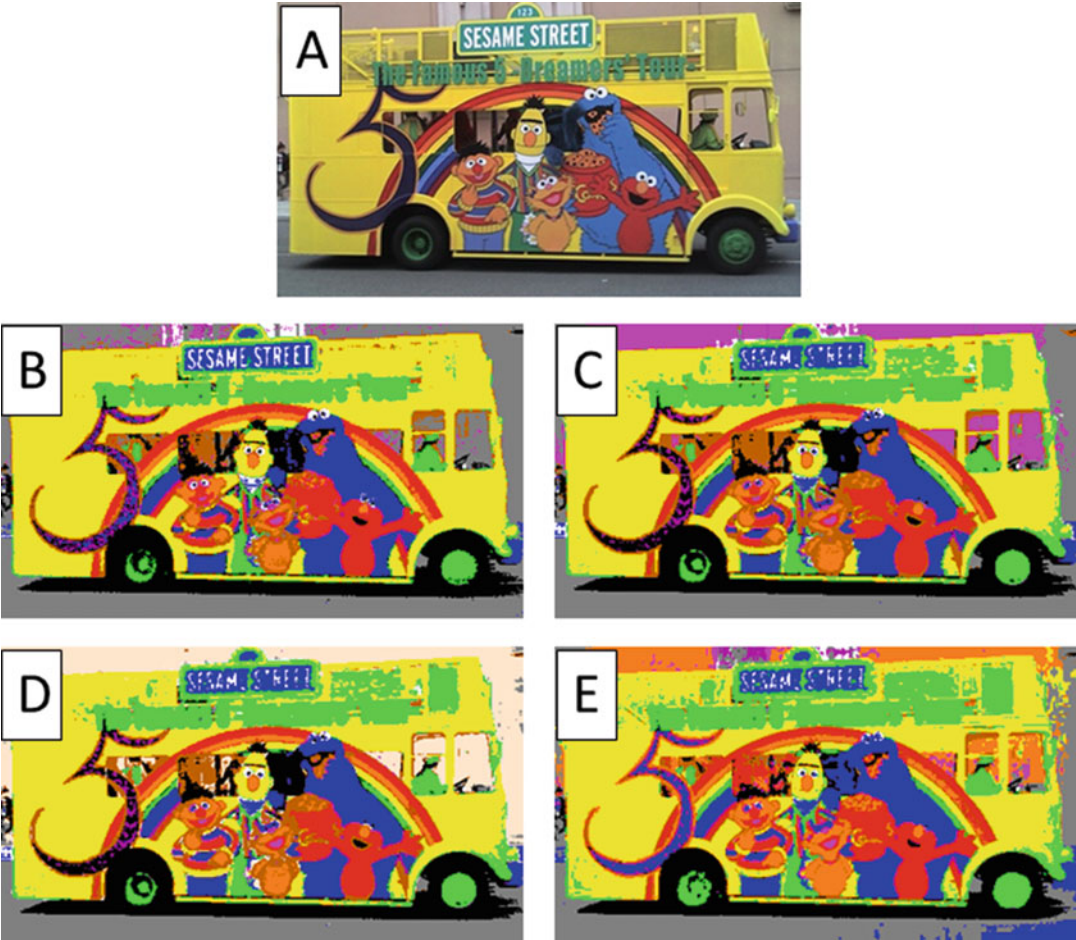
[12]; TSMES: Benavente et al.’s Triple Sigmoid-Elliptic Sigmoid model [15]; PLSA: van de Weijer et al.’s Probabilistic Latent Semantic Analysis [16] and NICE: Parraga and Akbarinia’s Neural Isoresponse Color Ellipsoids [6]. The data for LGM, MES, TSM, SFKM, and TSEM was obtained from Table 2 in Benavente et al. [15]

	Berlin and Kay results			Sturges and Whitfield results		
	Coincidences	Errors	% errors	Coincidences	Errors	% errors
LGM	161	49	23.33	92	19	17.12
MES	182	28	13.33	107	4	3.6
TSM	185	25	11.9	108	3	2.7
SFKM	193	17	8.1	111	0	0
TSEM	193	17	8.1	111	0	0
PLSA	187	23	12.3	109	2	1.8
NICE	206	4	1.9	111	0	0

embraced this idea to model color pixels (words) in images (documents) through a combination of color names (topics). Akbarinia and Parraga [20] showed that the same set of images from [16] can be used to learn the parameters of their previously proposed ellipsoidal model, and furthermore to extend it to new color names. Figure 2 shows a comparison of some of these results.

Results obtained by some of the computational color naming models mentioned above when applied to the image in panel A. Panels B, C, D, and E correspond to PLSA [16], NICE [6] (with 11 color terms), NICE [20] (with 12 color terms), and TSMES [15] respectively.

Color names have not yet been faithfully modeled by convolutional neural networks (CNNs), mainly because deep learning requires a large dataset for its training, which currently does not exist for color names. To date, there have been few attempts to create such datasets. Cheng et al. [21] used VGG16 to learn the color names of pedestrian’s clothing items. Despite achieving a better performance in their own introduced fashion dataset, the resulting network lacks generalization to images of other environments. In summary, despite being a promising technological development, the use of CNNs to produce a more generic color naming model remains an open challenge.



Color Name Applications in Computer Vision, Fig. 2 Results obtained by some of the computational color naming models mentioned above when applied to

the image in panel A. Panels B, C D and E correspond to PLSA [16], NICE [6] (with 11 color terms), NICE [20] (with 12 color terms) and TSMES [15] respectively

Applications

Color names are actively used as a rich source of feature descriptors in a wide range of classical computer vision applications [19] that can be broadly categorized as: classification, recognition, tracking, and retrieval.

Classification: A comparison of two sets of feature descriptors, one based on the classical eleven color names and another on raw hues, suggests that adding color names increases the robustness of the well-known bag-of-words method in the task of image classification [22]. Furthermore, the fusion of standard computer vision feature descriptors, e.g., HOG, with

color names, has been reported to improve the performance of object detection [23]. Overall, it can be argued that an explicit representation of color names is an effective method for improving object detection and classification.

Recognition: A similar fusion is proved to be successful in recognition tasks as well. Color names have been reported to provide a greater balance of photometric invariance and discriminative power. These properties make color names a suitable aid for action recognition [24]. Also, it has been shown that a feature representation grounded on 16 salient color terms facilitates person recognition and reidentification in video surveillance datasets [25].

Tracking: When it comes to visual tracking, it is well established that chromatic information provides a significant aid to the task; however, it is computationally demanding. This complexity can be overcome by using color names, which have been shown to have the efficiency of raw chromatic information, while being computationally more affordable [26].

Retrieval: color names are one of the first attributes chosen by humans when describing an object. With this in mind, Zheng et al. [27] supplemented a SIFT feature with basic color names, boosting the task of image retrieval by reducing false positives and improving query time. In order to reduce the semantic gap in algorithms of image retrieval, Liu et al. [28] proposed to associate each region of an image with a color name. This technique better meets user expectations when they use color names to describe their query.

A recent online survey investigated which color names are commonly used without restrictions of choice [29]. The results suggest that average individuals use more than 11 color names. Based on these results, Yu et al. [18] demonstrated that including a larger number of color names (28), instead of the traditional 11, increases the discriminative power of the color descriptors in the tasks discussed above. Griffin and Mylonas [30] used the data from [29] as a color metric to geometrically categorize the color space. Interestingly, the number of emerging regions within the RGB cube corresponds well to the optimum names used in [18].

During the past decade, the state of the art in the above applications has shifted from classical computer vision algorithms to deep learning approaches where the optimization of the network's parameters (millions of them) is often best achieved in an end-to-end training procedure. In other words, the features learnt by these networks are mainly data-driven. In most cases, imposing tailor-made features is likely to complicate the algorithms and impoverish the results. For this reason, incorporating color names to otherwise "pure" deep-learning algorithms has been little explored.

Outline for Future Research

Although much has been done in the past, there is still large room for improvement before achieving human-like performance in color-naming applications. The main problem is that humans are "noisy" when it comes to categorizing objects by color, and this noise has both perceptual and cognitive sources. Another great limitation has to do with statically assigning a color term to a given RGB value. In this paradigm, an arbitrary RGB value will always have the same color name regardless of its context and surrounding. Very few works (see [20]) have tried to address this point by dynamically adjusting the categorical regions to the image context. Given that human color perception is strongly influenced by local and global context, this dynamic color naming approach seems promising. On the other hand, modern CNN models also promise to advance the field by expanding the number of categories and learning some of the cultural variability; however, now there is a crippling lack of reliable datasets, although this is likely to improve in the future.

As a corollary, it is worth mentioning that the rewards for creating a robust color-naming paradigm are enormous, since many important fields (e.g., social robotics, image understanding, etc.) rely on it. There are also important advances to be made in helping visually impaired people appreciate and navigate a colorful world.

Cross-References

- ▶ [Ancient Color Terminology](#)
- ▶ [Berlin and Kay Theory](#)
- ▶ [Color Categorical Perception](#)
- ▶ [Cross-cultural Communication](#)
- ▶ [Dynamics of Color Category Formation and Boundaries](#)
- ▶ [Effects of Color Terms on Color Perception and Cognition](#)
- ▶ [Infant Color Categories](#)
- ▶ [Unique Hues](#)
- ▶ [Vantage Theory of Color](#)
- ▶ [World Color Survey](#)

References

1. Brainard, D.H., Hurlbert, A.C.: Colour vision: understanding #thedress. *Curr. Biol.* **25**, R551–R554 (2015)
2. Berlin, B., Kay, P.: *Basic Color Terms: their Universality and Evolution*. University of California Press, Berkeley/Oxford (1991)
3. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
4. Kay, P., McDaniel, C.K.: The linguistic significance of the meanings of basic color terms. *Language*. **54**(3), 610–646 (1978)
5. Olkkonen, M., Hansen, T., Gegenfurtner, K.R.: Categorical color constancy for simulated surfaces. *J. Vis.* **9**: 6 1–18 (2009)
6. Parraga, C.A., Akbarinia, A.: NICE: a computational solution to close the gap from colour perception to colour categorization. *PLoS One*. **11**, e0149538 (2016)
7. Maerz, A.J., Paul, M.R.: *A Dictionary of Color*. McGraw-Hill Book Company, inc., New York (1930)
8. Lammens, J.M.G.: A computational model of color perception and color naming (Ph.D. thesis). In: *Computer Science Department*, p. 253. State University of New York, Buffalo (1994)
9. Benavente, R., Vanrell, M., Baldrich, R.: Estimation of fuzzy sets for computational colour categorization. *Color Res. Appl.* **29**, 342–353 (2004) (Download model here)
10. Sturges, J., Whitfield, T.W.A.: Locating basic colors in the Munsell space. *Color Res. Appl.* **20**, 364–376.2 (1995)
11. Mojsilovic, A.: A computational model for color naming and describing color composition of images. *IEEE Trans. Image Process.* **14**, 690–699 (2005)
12. Seaborn, M., Hepplewhite, L., Stonham, J.: Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recogn.* **38**, 165–177 (2005)
13. Menegaz, G., Troter, A.L., Sequeira, J., Boi, J.M.: A discrete model for color naming. *EURASIP J. Appl. Signal Process.* **2007**, 113–113 (2007)
14. MacLaury, R.E.: From brightness to hue - an explanatory model of color-category evolution. *Curr. Anthropol.* **33**, 137–186 (1992)
15. Benavente, R., Vanrell, M., Baldrich, R.: Parametric fuzzy sets for automatic color naming. *J. Opt. Soc. Am. A*. **25**, 2582–2593 (2008)
16. Joost Reference Here
17. Yendrikhovskij, S.N.: A computational model of colour categorization. *Color Res. Appl.* **26**, S235–S238 (2001)
18. Yu, L., Zhang, L., van de Weijer, J., Khan, F.S., Cheng, Y., Parraga, C.A.: Beyond eleven color names for image understanding. *Mach. Vis. Appl.* **29**, 361–373 (2018)
19. Van De Weijer, J., Khan, F.S.: An overview of color name applications in computer vision. In: *International Workshop on Computational Color Imaging*, Mar 2015, pp. 16–22. Springer, Cham (2015)
20. Akbarinia, A., Parraga, C.A.: Colour terms: a categorisation model inspired by visual cortex neurons. *arXiv preprint arXiv:1709.06300* (2017)
21. Cheng, Z., Li, X., Loy, C.C.: Pedestrian color naming via convolutional neural network. In: *Asian Conference on Computer Vision*, Nov 2016, pp. 35–51. Springer, Cham (2016)
22. Khan, F.S., Van de Weijer, J., Vanrell, M.: Modulating shape features by color attention for object recognition. *Int. J. Comput. Vis.* **98**(1), 49–64 (2012)
23. Khan, F.S., Anwer, R.M., Van De Weijer, J., Bagdanov, A.D., Vanrell, M., Lopez, A.M.: Color attributes for object detection. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3306–3313. IEEE (2012)
24. Khan, F.S., Anwer, R.M., Van De Weijer, J., Bagdanov, A.D., Lopez, A.M., Felsberg, M.: Coloring action recognition in still images. *Int. J. Comput. Vis.* **105**(3), 205–221 (2013)
25. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.Z.: Salient color names for person re-identification. In: *European Conference on Computer Vision*, Sept 2014, pp. 536–551. Springer, Cham (2014)
26. Danelljan, M., Shahbaz Khan, F., Felsberg, M., Van de Weijer, J.: Adaptive color attributes for real-time visual tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1090–1097 (2014)
27. Zheng, L., Wang, S., Liu, Z., Tian, Q.: Packing and padding: coupled multi-index for accurate image retrieval. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1939–1946 (2014)
28. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: Region-based image retrieval with high-level semantic color names. In: *11th International Multimedia Modelling Conference*, pp. 180–187. IEEE (2005)
29. Mylonas, D., MacDonald, L.: Augmenting basic colour terms in English. *Color Res. Appl.* **41**(1), 32–42 (2016)
30. Griffin, L.D., Mylonas, D.: Categorical colour geometry. *PLoS One*. **14**(5), e0216296 (2019)

Color Names for Image Understanding

► Color Name Applications in Computer Vision

Color Naming

- [Chinese Color Language](#)
 - [Color Constancy and Categories](#)
-

Color Order Systems

José Luis Caivano¹ and Antal Nemcsics²

¹Secretaría de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

²Budapest, Hungary

Synonyms

[Color atlases](#); [Color models](#); [Color solids](#); [Color spaces](#)

Definition

Color order systems can be defined in three ways:

- (a) A system for categorizing colors. An arrangement of perceived colors, color stimuli, or material color samples according to certain rules.
- (b) A subset of the world of color according to three attributes that constitute the coordinates of the color system.
- (c) A rational plan for ordering and specifying all object colors by a set of material standards.

Color Systems: Review and Evaluation

The aim of introducing order into the vast set of colors that most humans are able to distinguish has existed since ancient times. Among the people who have proposed some kind of color order systems, there are famous philosophers,

architects, artists, scientists, and writers, most of whom are included in a recent book about the pioneers of color science [1]. We can mention Aristotle (c.350 BC), ► [Avicenna](#) (c.1000), Leon B. Alberti (1435), Leonardo da Vinci (1516), ► [Isaac Newton](#) (1704), Johann Heinrich Lambert (1772), and ► [Johann W. Goethe](#) (1810). In the twentieth century, other scientists and theorists, such as ► [Albert H. Munsell](#) (1905, 1907, 1915), ► [Wilhelm Ostwald](#) (1916, 1917), Arthur Pope (1922, 1929, 1949), Cándido Villalobos and Julio Villalobos-Domínguez (1947), Antal Nemcsics (1975), Harald Küppers (1978), and Frans Gerritsen (1987), just to mention a few, have been outstanding in formulating and building color order systems. This endeavor has also been pursued by organizations such as the Commission Internationale de l'Eclairage (1931, 1976), the Optical Society of America (1947–1977), the Swedish Standards Institution (1979), and others.

A color order system seeks to include samples of all colors in a spatial model, giving a specific position to each color and proposing some kind of logic that determines the whole organization. These models have very diverse shapes, according to the different authors and along times: linear scales, two-dimensional diagrams, chromatic circles, color triangles, and color solids of various shapes. Within this last type, different solutions have been introduced: cones, pyramids, double cones, double pyramids, spheres, and more or less irregular solids.

Old Linear Organizations and Diagrams

The most ancient color organizations are often simple lists of color names, linear scales – generally expressed in verbal form, without graphic representations – or two-dimensional diagrams in the form of color schemes, triangles or circles. Among the first group, two can be mentioned: the five basic colors of the old Chinese philosophy (grue [qing, a conflation of blue and green], red, yellow, white, and black), related to five elements (wood, fire, earth, metal, and water, respectively) and to the five metaphysical localizations (east,

south, center, west, and north, respectively), and the color scale of Aristotle (384 BC–322 BC), with white and black at the ends, and a series of intermediate chromatic colors linearly arranged between these two extremes of lightness and darkness.

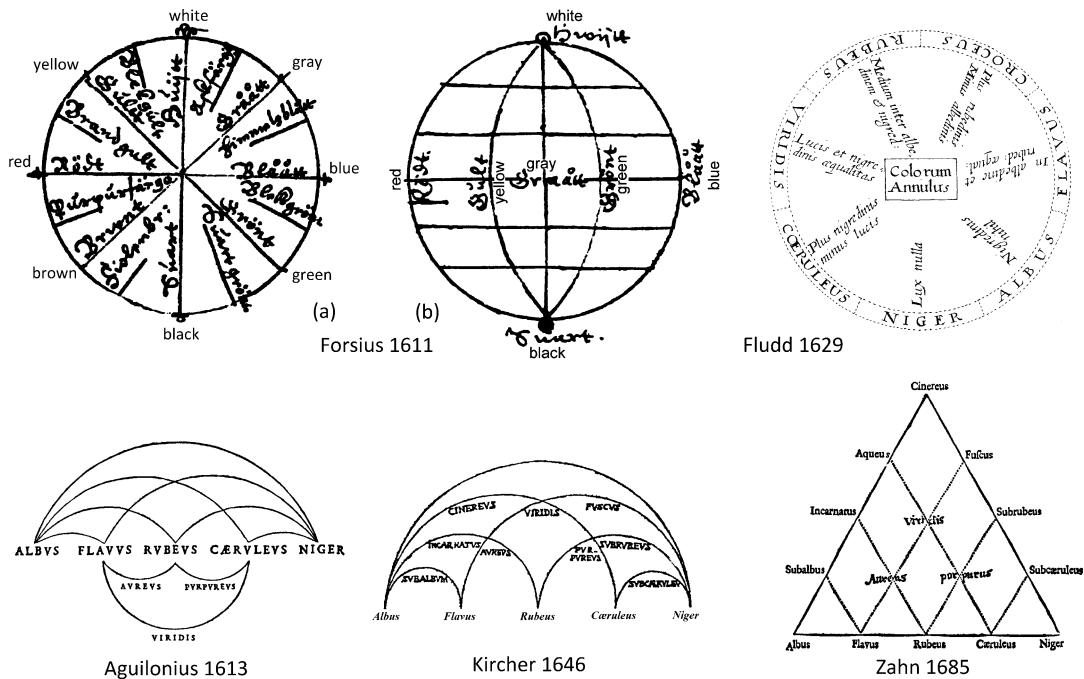
By the turn of the first millennium and during the next two centuries, there were some developments stemming on the Aristotelic school in Western Asia. ► **Avicenna** (980–1037), the Persian philosopher and scientist, expanded the linear sequence by establishing three possible paths from white to black. Another Persian polymath, ► **Nasir al-Din al-Tusi** (1201–1274) described a system of color mixture with five paths from white to black. By interpreting their texts and rendering them graphically, Kirchner and Bagheri obtained schemes that can be considered intermediate steps between linear arrangements and two-dimensional diagrams [2]. In parallel, a similar description is found in Europe around 1230 in the text *De colore* by the English scholar Robert Grosseteste (circa 1175–1253).

A verbal ordering of colors, which some later authors reconstructed as a chromatic circle or a square, and even as a double cone, a double pyramid, or a sphere [3], is found in the book *On painting* by the Italian architect Leon Battista Alberti (1404–1472).

A great deal of the *Treatise on painting* by Leonardo da Vinci (1452–1519), compiled after his death from his notebooks, is dedicated to insights about color. For the arrangement of colors, he proposed a simple scheme based on the oppositions black-white, blue-yellow, and green-red.

The Belgian Jesuit Franciscus Aguilonius [1] (1567–1617) published his *Opticorum libri sex* in 1613 in Antwerp, and the German Jesuit Athanasius Kircher (1602–1680) published his *Ars magna lucis et umbrae* in 1646 in Rome. Both texts include diagrams with a primary linear arrangement of colors connected by arcs that indicate the results of their mixtures (Fig. 1).

Aron Sigfrid Forsius [1] (1550–1624), a Finnish-born clergyman, cartographer, and astronomer who worked in Sweden, wrote a



Color Order Systems, Fig. 1 Object color diagrams of the seventeenth century

manuscript in 1611 where the first-known drawing of a ► **color circle** appears. Forsius' circle presents the following sequence: white, yellow, red, brown, black, green, blue, and gray, closing again in white (Fig. 1a). Forsius characterized this circle as "ancient" and drew an alternative for the arrangement of colors. Various theorists have interpreted this second diagram as a sphere, but this is controversial (Fig. 1b). If this can be demonstrated, it would be the first three-dimensional system in the history of color. Since Forsius did not write the word "sphere" in the accompanying text, and the drawing is schematic and not well fitted to the representation of a sphere in vertical projection or in perspective, it is difficult to accept this. In addition, the diagram has two misplaced colors, if we want to understand it as a logical natural color system: Blue and green should exchange places [4]. This is mere speculation, but since this diagram has led to a lot of contemporary discussions, interpretations, and misunderstandings, had it been more widely known in published form during Forsius time, instead of being kept as a manuscript in the Stockholm library, it is likely that its influence could have led other scholars to conceive a true color sphere in the seventeenth century, even through the way of misinterpreting Forsius diagram.

In 1629 Robert Fludd (1574–1637) published a circular diagram. However, rather than a two-dimensional circle, it is a color ring (*colorum annulus*), that is, a round sequence that includes black (*niger*) and white (*albus*) in the steps. It is a way of connecting the extremes of the Aristotelian linear scale by bending it.

In 1685, Johannes Zahn (1641–1701) published the first known color diagram in the form of a triangle (Fig. 1). But, again, it is not a triangle of full hues (as we understand it today), nor is it a triangular surface. It is a scheme of color relationships and mixtures, similar to Aguilonius' diagram, but using straight lines (instead of arcs) to connect the colors. That is, the colors are placed at the intersections, and they do not fill the surface. At the base of the triangle, we find a scale of three colors – yellow (*flavus*), red (*rubeus*), blue (*caeruleus*) – between white (*albus*) and black (*niger*). The mixture of white and black makes

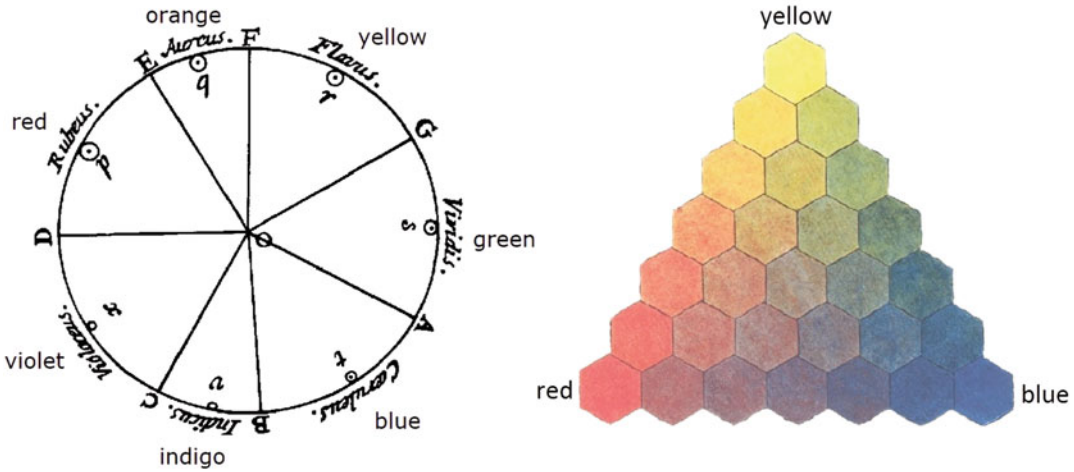
gray (*cinereus*), yellow and red make orange or golden (*aureus*), red and blue make purple (*purpureus*), and yellow and blue make green (*viridis*). There are the other mixtures as well.

These linear, circular, or triangular diagrams include black and white in their sequences and are not truly two-dimensional color spaces, in the modern sense.

Modern Two-Dimensional Color Arrangements

The first modern ► **color circle** appeared in 1704. ► **Sir Isaac Newton** (1642–1727) discriminated seven colors in the spectrum produced by the dispersion of light through a prism, and arranged them in the perimeter of a circle divided into seven portions, with the center occupied by the white light. The fact that Newton included seven colors (5 with a larger span interval and 2 with a shorter one) is due to the parallelism he made with the seven musical tones and half-tones of the diatonic scale. The circle is a representation of a mixture of spectral lights that blend in white at the center. In modern terms, it is an additive color mixture. Different from any previous circular representation, here the idea is that color mixtures fill out all the surface progressively, from the outer spectral sequence towards the white center (Fig. 2, left). However, the difference between additive and subtractive mixture was not clear at the time of Newton. It began to be understood only during the nineteenth century. Newton's circle is the direct antecedent, with some intermediate steps, of the development of the ► **CIE 1931** chromaticity diagram.

In 1708, very close to the publication date of Newton's *Optics*, a book on miniature painting was published, which at the end included two painted color circles. The authorship of these circles is uncertain; various scholars attribute them to Claude Boutet. The first circle has seven colors (like Newton), but the most interesting one is the second circle, which contains 12 hues, arranged in triads. This is the older known representation of the traditional color wheel used by painters until today, and also the first published color circle that



Color Order Systems, Fig. 2 Left: Newton's circle. Right: Mayer's triangle

represents a trichromatic mixture, with a triad of yellow, red, and blue, inscribed in an equilateral triangle, opposed to another triangle with green, orange, and purple. Other six intermediate hues complete the 12 hues circle. This circle with subtractive mixtures of red, yellow, and blue is a truly modern two-dimensional color diagram and is kept practically without changes until nowadays. We can see it adopted again by Moses Harris (1770), ► [Johann W. Goethe](#) (1810), ► [Philip Otto Runge](#) (1810), Charles Hayter (1826), George Field (1841), ► [Michel-Eugène Chevreul](#) (1864), Arthur Pope (1922), ► [Johannes Itten](#) (1961), and many others [1].

The first modern color triangle, with three chromatic hues, and with colors filling all the surface, was proposed in 1758 by ► [Tobias Mayer](#) (1723–1762), in a public lecture delivered in Göttingen, and published after Mayer's death by Georg Lichtenberg, in 1775. The triangle has red, yellow, and blue as ► [primary colors](#), and Mayer developed a color notation system and a theory of mixtures by quantifying and combining the amounts of each primary in terms of 12 steps. The method and theoretical formulas look as if they were following additive principles, but we have to remember that the difference between additive light mixture and subtractive pigmentary mixture was not well understood at that time, and

thus, the system is somehow confusing in this sense. The colored representation of the triangle was produced with pigments (Fig. 2, right), and a dull gray or brown color is obtained by the mixture of the three primaries, instead of white that would result from an additive mixture, as in the case of Newton.

Mayer's unpublished papers were collected posthumously in a book edited by Georg Lichtenberg, including a hand-colored version of the triangle. Even when Mayer did not produce a drawing of a three-dimensional model, it can be interpreted from his formulation. In a commentary to the publication, Lichtenberg describes a prism that can be derived from Mayer conception and states that the result “can be reached more elegantly with two identical pyramids having a common base.” This description has been rendered graphically by Heiwing Lang [5].

Two Important Three-Dimensional Systems Before the Twentieth Century

Around 1760, ► [Johann Heinrich Lambert](#) (1728–1777) was aware of reports of Mayer's triangular system published in Göttingen. Lambert devised a color pyramid, which was published in 1772. This was the first color solid

expressly developed and drawn as such by its author. The pyramid has a triangle with scarlet (equivalent to primary red), amber (equivalent to primary yellow), and blue at its base, from whose mixtures Lambert was able to obtain black in the center of the triangle at the base. Hence, this is an arrangement of subtractive mixture of pigments. By placing white in the upper vortex, the system is completed. Along the vertical axis, we find successive triangles that become lighter and lighter from the base to the top of the pyramid.

The German painter ► [Philip Otto Runge](#) (1777–1810) built a color sphere that is considered the ancestor of most of the twentieth-century color systems and could be termed “modern” for this. Runge’s sphere, published in 1810, is the significant result of a sustained evolutionary process initiated in the Renaissance. The chromatic circle, at the equator of the sphere, is arranged on the basis of the red-yellow-blue triad of ► [primary colors](#), plus the green-orange-violet triad of secondary colors, as in Boutet and Goethe’s circles.

Classification of Color Order Systems

Color systems can be categorized into two large groups: color stimulus systems and color perception systems. Color stimulus systems are established in the twentieth and twenty-first centuries. Their developers are almost without exception physicists and information specialists. They are based on systematizing radiation energy of different wavelengths creating the experience of different colors. Some of these systems are registered as industrial international standards. Their detailed description may be found in chapters of the Encyclopedia dealing with the measurement of color and computerized display of color.

The roots of the establishment of color perception systems appear in the distant past. Their creators belong to different professions and include among others, painter artists, architects, writers, poets, philosophers, doctors, priests, bishops, chemists, botanists, and many other crafts, representing the diversified world of colors.

The established multitude of perception-based color order systems can be categorized into eight groups in terms of aims, classification methods, and definition of color patterns by color notation. Irrespective of their categorization, they possess numerous identical attributes.

Every perception-based color system possesses a color atlas containing a set of color samples. Early on, colors were placed on plane surfaces, within regions represented by a triangle, square, or circle. Later, in the eighteenth century, it turned out that the multitude of colors could only be arranged within a three-dimensional space. With the exception of a few systems (Munsell, and Coloroid, among them), colors were arranged on the surface of, or within, regular bodies. The most frequently applied bodies were the following: a triangle-based pyramid (Lambert and Mayer), a circle-based double cone (Ostwald [6]), a square-based double cone (Höfler), a circle-based truncated cone (Baumann-Prase), a cube (Hickethier), a sphere (Runge), a half-sphere (Chevreul), a body of rotation consisting of cones (Wilson), a circle-based cylinder (Frieling), and a body constructed of cube-octahedrons (OSA).

The irregular shapes of the color bodies of Munsell and Coloroid systems are the result of the fact that, as opposed to other color systems, they take into account the magnitude of chromatic strength and lightness of colors of different hues, as well as practical issues pertaining to the production of colors of high saturation and high luminance for some hues.

All perception-based color order systems are discontinuous, the single exception being the Coloroid system. They contain only a definite number of colors from the color space. In most of the systems, the letter or number notating the color indicates the location of the color in the color atlas. In most of the systems, the color notation gives the additive or subtractive components for mixing the given color. In some systems, it defines a particular color perception. In the case of some color systems (e.g., Munsell, NCS), the color coordinates of the CIEXYZ color system are provided for a definite number of their colors. This allows the calculation of CIE coordinates –

by interpolations with different degrees of accuracy – also for colors not presented in the color atlas. A direct transformation relation exists only between Coloroid and CIEXYZ which enables defining with equal accuracy in both systems all colors distinguishable by the human eye, that is, more than a million colors.

Review List of Ancient and Modern Color Order Systems

Systems of Color Stimuli

One criterion for classifying systems of color stimuli is whether they depend on some particular device to produce them or not. In a device-dependent color system, the resultant color depends on the equipment used and its calibration. The same RGB stimulus generated in a computer display (for instance, $R = 255$, $G = 0$, $B = 255$, which yields a magenta) will give different results depending on the type of display used (CRT, LCD, LED), its brand, the brightness and contrast setup, etc. In a device independent color system, on the other hand, the coordinates that specify a color will indicate the same color, independently of the medium by which it is produced.

Device-Dependent Systems

RGB 1983, HLS 1983, HSV 1983, CMY 1983, CMYK 1983, Adobe RGB 1998, SRGB, HKS, HSV.

Device-Independent Systems

Schrödinger 1920, Luther 1924, Nyberg 1928, Rösch 1928, Wright 1928, Guild 1931, ► [CIE 1931](#), MacAdam 1942, Stiles 1946, Brown 1951, CIE 1960, Wyszecki 1963, CIE 1964, Judd 1963, ► [CIELAB 1976](#), CIELUV 1976, CIECAM 1997, ► [CIECAM 2002](#).

Systems of Color Sensations

Systems Based on Experience

Grosseteste 1230, Alberti 1435, Leonardo da Vinci 1516, Forsius 1611, Aguilonius 1613, Fludd 1629, Boutet 1708, Harris 1766,

Schiffmüller 1772, Baumgärtner 1803, Runge 1810, Hay 1828, Merimée 1839, Field 1841, Barnard 1855, Delacroix 1856, Jännicke 1878, Henry 1889, Kreutzer 1894, Pope 1922, Becke 1924.

Systems Bearing a Parallelism to Music

Newton 1704, Adams 1862, Seurat 1887.

Systems Based on Additive Color Mixing

Newton 1704, Wunsch 1792, Chevreul 1839, Doppler 1847, Grassmann 1853, Maxwell 1855, Helmholtz 1860, Bezold 1874, Ridgway 1913, Ostwald 1916, ► [CIE 1931](#), Baumann-Prase 1941, Jacobson 1948, Rabkin 1950.

Systems Based on Subtractive Color Mixing

Kircher 1646, Waller 1686, Mayer 1758, Lambert 1772, Sowerby 1809, Hayter 1826, Benson 1868, Hering 1878, Rood 1879, Lovibond 1887, Lacouture 1890, Höfler 1897, Plochère 1946, Colorizer 1947, Müller 1948.

Printed Screen Systems

Wilson 1942, Villalobos 1947, Hickethier 1963, Küppers 1978.

Educational Systems Based on Psychological Effects

Goethe 1810, Schopenhauer 1830, Wundt 1874, Ebbinghaus 1902, Klee 1922, Itten 1923, Boring 1929, Birren 1934, Frieling 1968, CMM Silvestrini 1986, Gerritsen 1987, Albert-Vanel 1990.

Perceptually Equidistant Color Systems

Munsell 1915, Johansson 1937, DIN-Richter 1953, Hesselgreen 1953, OSA-Wyszecki 1960, Manfred Adam 1966, Kobayashi 1966, Coloroid-Nemesics 1975, Acoat Color Codification, NCS (Hård, Sivik, Tonnquist) 1979, Eusemann 1979, Chromaton 1981.

Practical Color Collections

Maerz 1930, Jeanneret (Le Corbusier) 1931, ISCC-Kelly 1955, Gericke-Schöne 1969, Chroma Cosmos 1987, Master Atlas 1988, Cler 1992, RAL 1993, Pantone, EuroColor 1984.

Color Systems Used Today, Recommended for Art and Design

Munsell Color System

The foundations for the most up-to-date idea of color systems were first laid down by ► [Munsell](#), who published his *Book of color* [7] and the pertinent color sample collection in 1915 for the first time. This system was further developed in 1943 and correlated to the ► [CIE 1931](#) colorimetry system. In 1956, it was extended to include very dark colors. Later, the Munsell Color Company was founded for re-editing at regular intervals, in the original quality, the color collection of the system. Since 1979, it is also published in Japan under the title *Chroma Cosmos 5000*.

The Munsell system is still one of the most popular color systems. Its codes are up to this day the most common color identification numbers in the international literature. Colors are identified by three kinds of perceptual data: hue H , value V (lightness), and chroma C . These data are the three coordinates of the Munsell color solid, characterized by cylindrical coordinates and corresponding to the three characteristics of visual perception.

The hue scale of the Munsell system is divided into 100 perceptually equal parts according to 10 shades each, of the following five reference and five mixed colors: red (R), yellow-red (YR), yellow (Y), green-yellow (GY), green (G), blue-green (BG), blue (B), purple-blue (PB), purple (P), and red-purple (RP) (Fig. 3).

The vertical axis of the Munsell color solid accommodates gray (neutral, N) colors. The achromatic scale is divided in a perceptually equidistant manner from 0 to 10. Munsell value (lightness) of a perfectly absorbing surface (ideal black) is 0, while that of a perfectly and diffusely reflecting surface (ideal white) is 10. Munsell chroma increases with the distance from the achromatic axis (Fig. 4). In the Munsell system, colors are marked as $H\ V/C$, for example, 10RP 5/6. The shape of the Munsell color solid is shown in Fig. 5. The relationship between the Munsell color system and the CIE colorimetric system has been tabulated.

Natural Color System (NCS)

One of the latest ideas on color systematization relies on the Hering-Johansson theory, materialized as the color atlas by Hesselgren issued in 1953. Based on this atlas, in 1972 Hård, Sivik, and Tonnquist developed the Natural Color System (NCS), which was adopted as a Swedish standard in 1979.

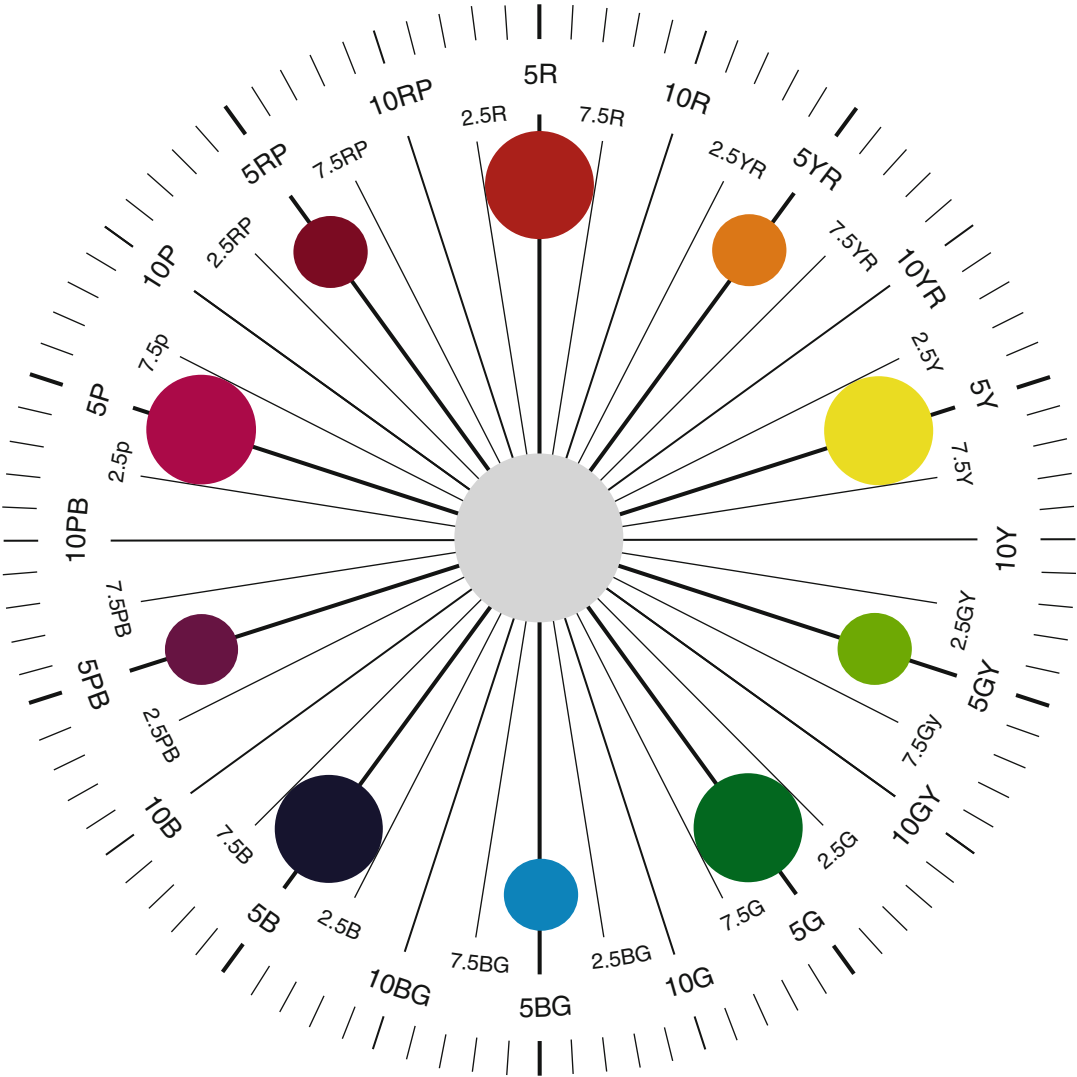
The authors of this system started from ► [Ewald Hering's](#) idea about six elementary color perceptions, that is, white (W), black (S), yellow (Y), red (R), blue (B), and green (G). In the NCS, every color is described by the degree of its similarity to the six elementary colors [8].

A color cannot be similar to more than two hues. Yellow and blue exclude each other, and so do red and green. The sum of the color variables defines the NCS chromaticness (c) of the perceived color and their ratio its hue (Φ).

In the NCS, colors are marked as $sc\text{-}\Phi$, for instance, 2040-G40Y, where s is blackness, c is chromaticness (a magnitude associated with saturation), and Φ is the hue of the color. In the color notation, the s and c values – both of two digits – are written without space and separated from Φ by a hyphen. The hue of the color in the example above is intermediate between green (G) and yellow (Y) in a ratio of 40 to 60.

The geometric form that the NCS assumes is a regular symmetric double cone with white and black at the vertices, while the other four preferential hues are on the circle of full colors, at corners of a square touching this circle (Figs. 6, 7, and 8). This color solid is of the same shape as that of the Ostwald system [6]. Also, the basic principle of some variables was adopted from the Ostwald system, but with a modified definition and scale. Similarities and differences between the CIE and NCS color spaces have been tabulated, but no exact mathematical correlations were established.

The significance of the NCS is that it operates with a tetrachromatic color description as a possible way of visual color description.



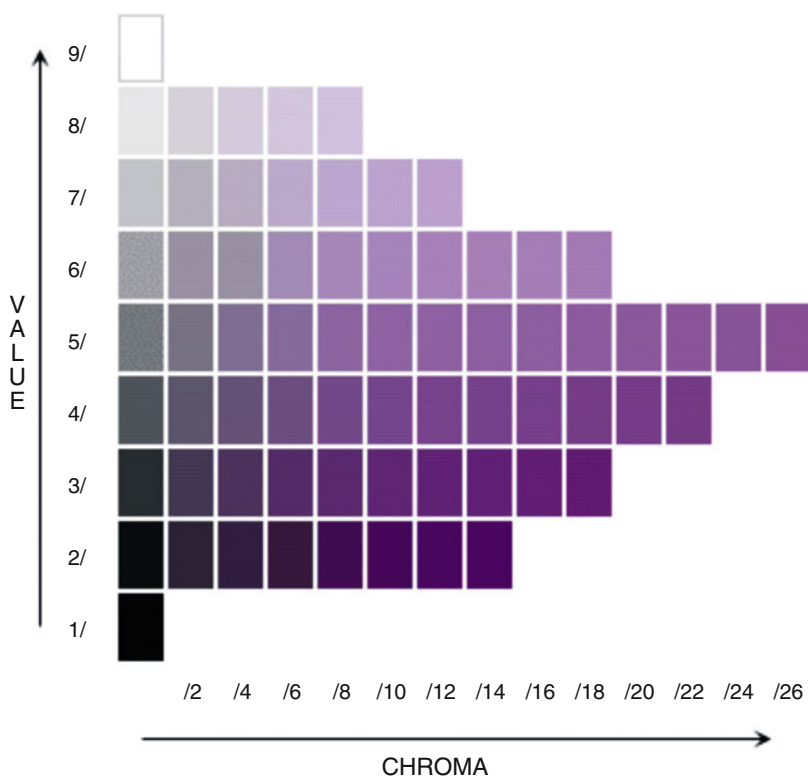
Color Order Systems, Fig. 3 Hue circle of the Munsell system

Coloroid System

Coloroid is a system of surface colors illuminated by daylight and sensed by normal color vision observers, built on harmonic color differences according to sensation, that well approximates aesthetic uniformness. Coloroid has a direct transformation relation with the CIEXYZ system. It was elaborated at the Budapest Polytechnic University by Nemcsics and first published in 1975. Since 1982 it has been a Hungarian technical

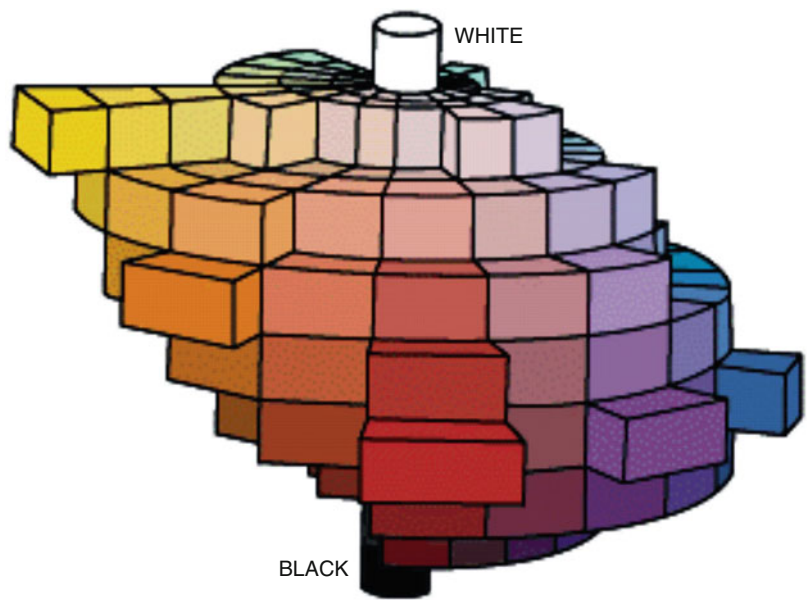
directive, and it became a Hungarian standard in 2002 [9–11].

Coloroid coordinates are semipolar coordinates, representing the members of the population of colors placed inside a cylinder, to be used for the explicit definition of color points, namely, the angular coordinate representing numerically the hue of the color (A), the radial coordinate representing numerically the saturation of the color (T), and the vertical axial coordinate representing the luminosity of the color (V).



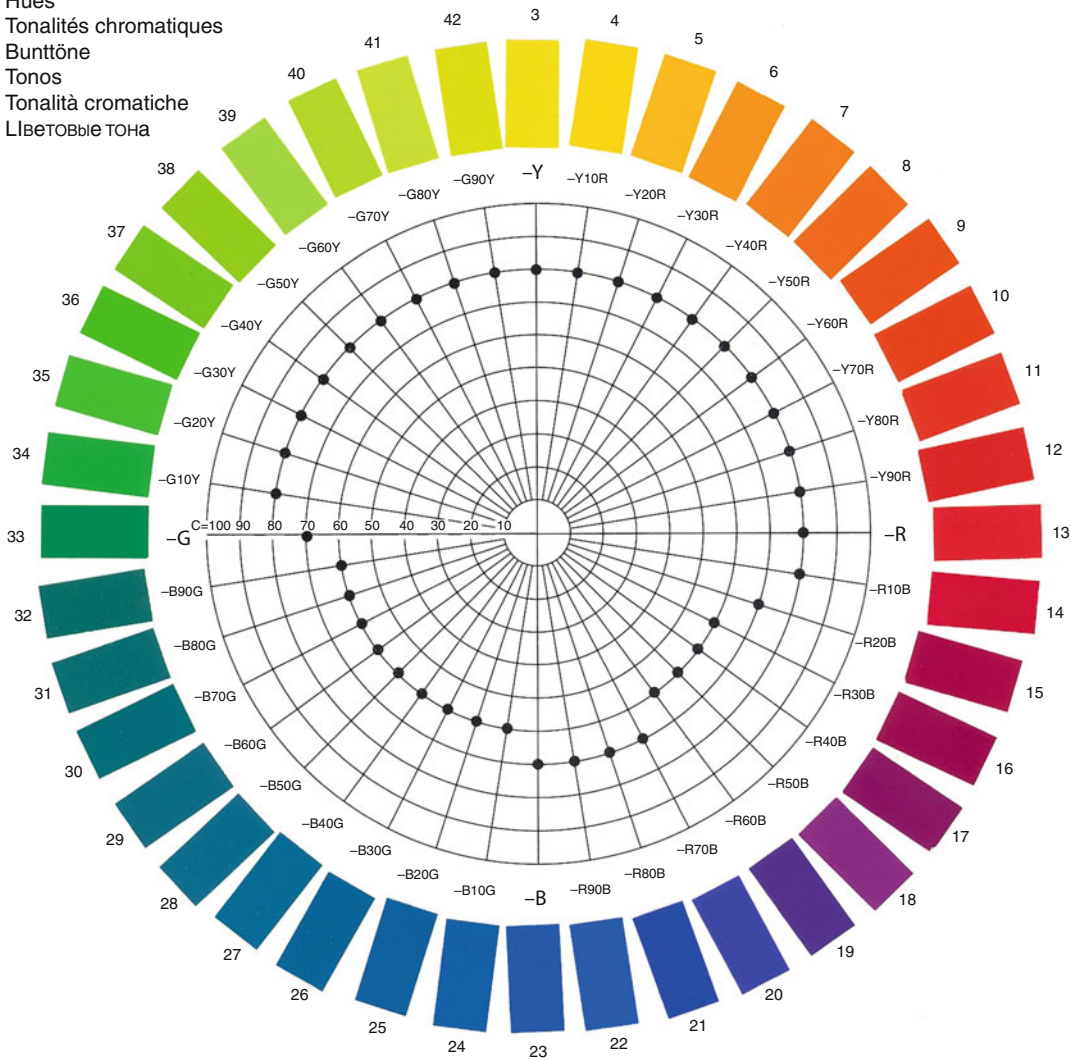
Color Order Systems, Fig. 4 Axial section of the Munsell color solid, displaying a certain hue with the related chroma and value scales

Color Order Systems, Fig. 5 Color solid of the Munsell system





Hues
Tonalités chromatiques
Bunttöne
Tonos
Tonalità cromatiche
Цветовые тона

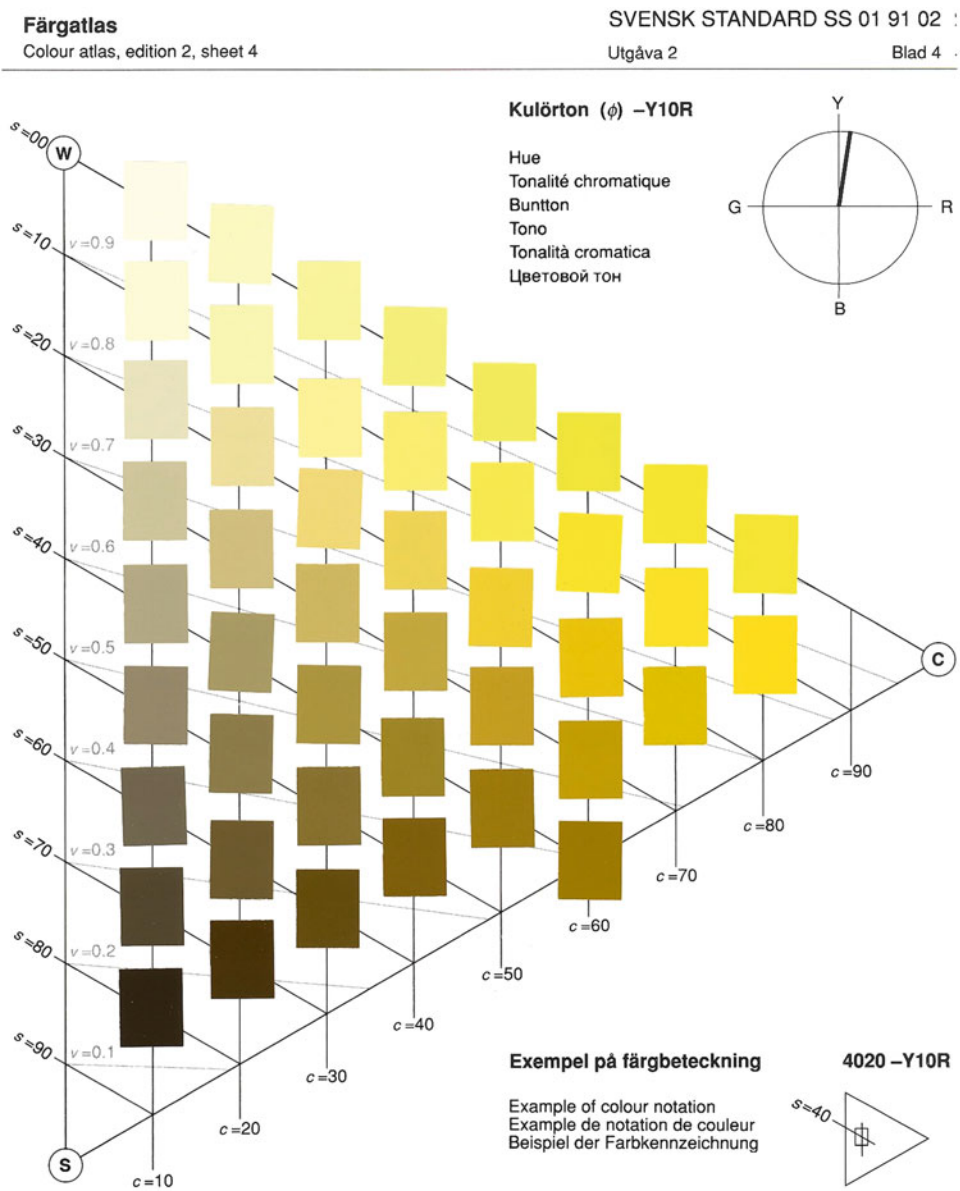


Color Order Systems, Fig. 6 Hue circle of the Natural Color System

Absolute white (W) is placed on the upper limit point of the axis of the color space. It is the color of a surface illuminated by the CIE D65 illuminant, with perfectly scattered reflection, having both Coloroid luminosity value and Yw color component value of 100. Absolute black (S) is placed on the lower limit point of the axis of the color space. It is the color of a surface illuminated by the CIE D65 illuminant, perfectly light absorbing ($\beta = 0$ luminance factor), having both

Coloroid luminosity value and Ys color component value of 0.

The Coloroid limit colors are the most saturated colors that can be drawn onto the surface of the cylinder comprising the Coloroid space, located along a closed curve. The Coloroid basic colors are 48 different limit colors characterized with integer numbers, being located at approximately identical number of harmony intervals to each other. The Coloroid basic colors are recorded

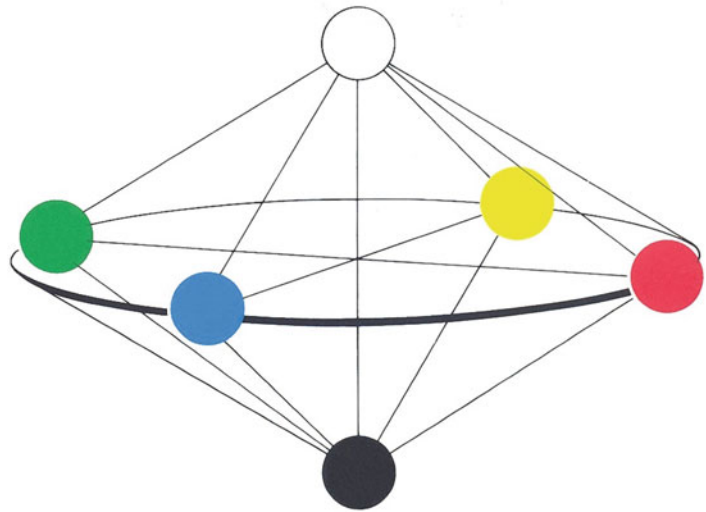


Color Order Systems, Fig. 7 Axial section of the NCS solid, displaying the Y10R hue with the variation of darkness and chromaticness

in the CIE 1931 diagram by the jangle. The jangle is the angle of the half line originated from the D65 point of the CIE 1931 chromaticity diagram to the x axis.

The Coloroid color planes are the half planes delimited by the achromatic axis of the color space, having the same hue and dominant

wavelength. In each color plane, colors are enclosed by the neutral axis and two delimiting curves. The shape of surfaces enclosed by delimiting curves is different for each hue and depends on the luminosity of the color located on one apex of the color plane. Along the vertical lines of the nets drawn on the color planes,

Color Order Systems,**Fig. 8** Schematic representation of the solid of the Natural Color System

Coloroid saturation values are identical; along their horizontal lines, Coloroid luminosity values are identical. Colors implemented with various means or colors created in nature, belonging to individual color planes, are enclosed by internal delimiting curves.

The Coloroid basic hues are the hues belonging to Coloroid basic colors. Similarly to basic colors, there are 48 Coloroid basic hues. In color planes A10 to A16 yellow, in color planes A20 to A26 orange, in color planes A30 to A35 red, in color planes A40 to A46 purple and violet, in color planes A50 to A56 blue, in color planes A60 to A66 cold green, in color planes A70 to A76 warm green hued colors exist.

Coloroid saturation is a characteristic feature of the surface color quantifying its saturation, that is, its distance from the color of the same Coloroid achromatic luminosity measured on a scale that is aesthetically near to uniform. Its denotation is T. The saturation of the limit colors is equal to 100. The saturation of absolute white, absolute black, and gray is equal to 0. In the Coloroid space, colors of identical saturation are located equidistant to the achromatic axis, on a coaxial cylinder.

Coloroid lightness is a characteristic feature of the surface color denoting the distance measured from absolute black on an aesthetically near-uniformly graduated scale. Its denotation is

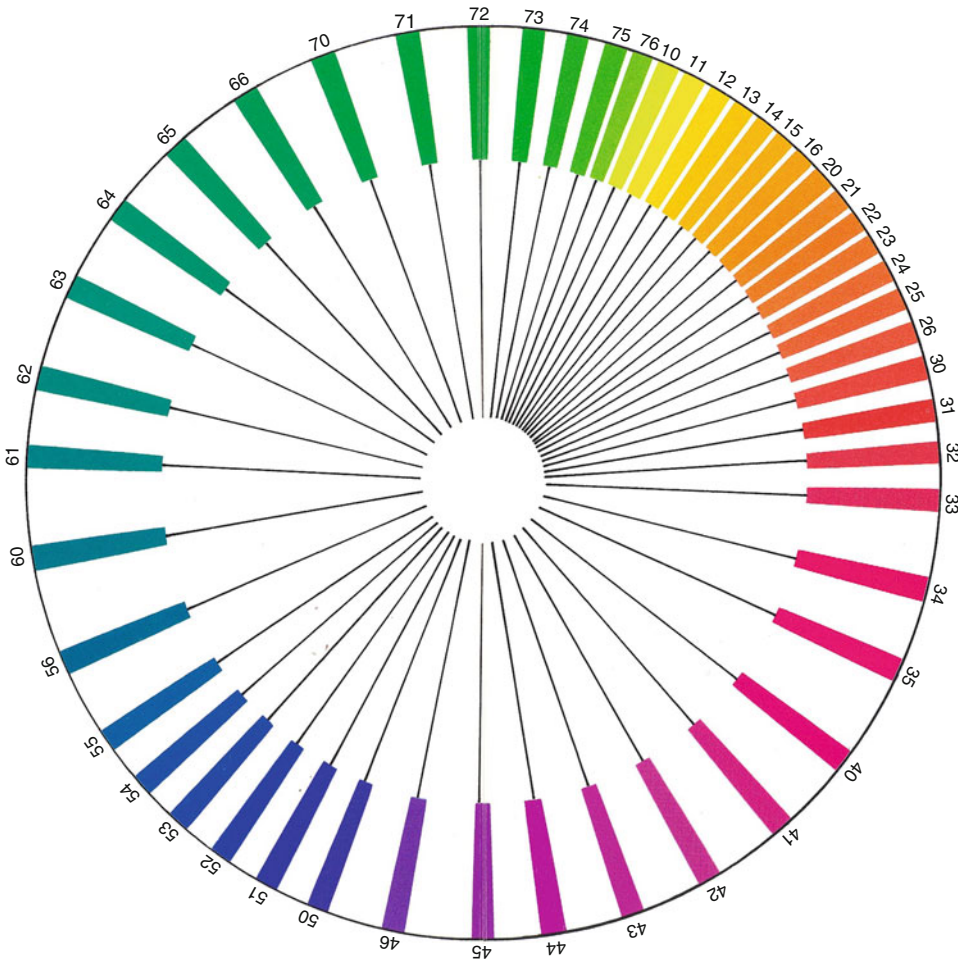
V. The lightness of absolute black is equal to 0. The lightness of absolute white is 100. In the Coloroid space, colors of identical lightness are located in planes perpendicular to the achromatic axis. Numerical values of the Coloroid lightness of a surface color are determined by the expression $V = Y^{1/2}$.

Coloroid hue is a characteristic feature of the surface color denoting its hue on a scale distributed into 48 sections on an aesthetically near-uniformly graduated scale. Its denotation is A. The Coloroid hue of the surface color is a function of the dominant wavelength. In a Coloroid space, surface colors having identical hues lie in the Coloroid color planes.

The color notation in the Coloroid system is expressed by hue (A)–saturation (T)–lightness (V), for instance, 13–22–56 (Figs. 9, 10, and 11).

Küppers' Atlas and Rhombohedric Color System

The German engineer Harald Küppers published an atlas specifically useful for the graphic arts and the printing industry, containing more than 5,500 nuances [12]. The color samples of this atlas were produced by the technique of four-color printing, mixing the four transparent dyes, yellow,



Color Order Systems, Fig. 9 Hue circle of the Coloroid system

magenta, cyan, and black, with the concurrence of the white background of the paper.

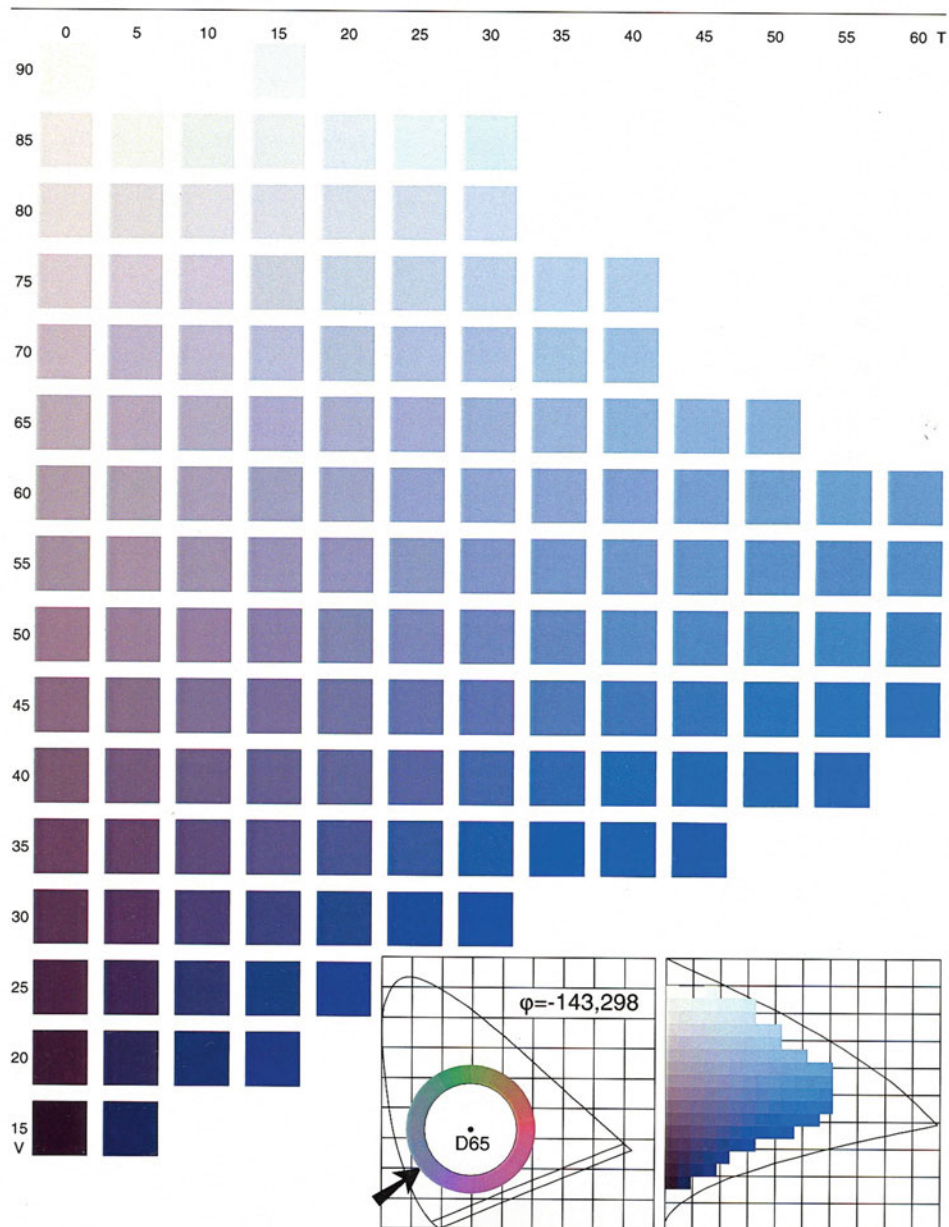
The gradations of the mixtures are expressed in percentages directly equivalent to the proportion of surface covered by each dye, so that this notation not only is useful to designate the different nuances but also as a formula to produce the colors. The variation is made by differences of 10% between an individual sample and the next one and between a chart of colors and the next one.

The published color charts are divided into five series; three of them are said to be of achromatic mixture because of the intervention of

black, and the other two are considered of chromatic mixture due to the exclusive intervention of yellow, magenta, and cyan. In the first three series, black is added to a mixture of two chromatic dyes – made in 10% steps of variation and held constant for the whole series – in 10% steps of variation from one chart to another. In the fourth series, yellow is added in successive charts to a fixed mixture of two chromatic dyes (magenta and cyan). Series 1–4 have 11 charts, with 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 99% of the dye being added to the fixed binary mixture. The fifth series has two additional charts: one with

Coloroid színatlasz

A54



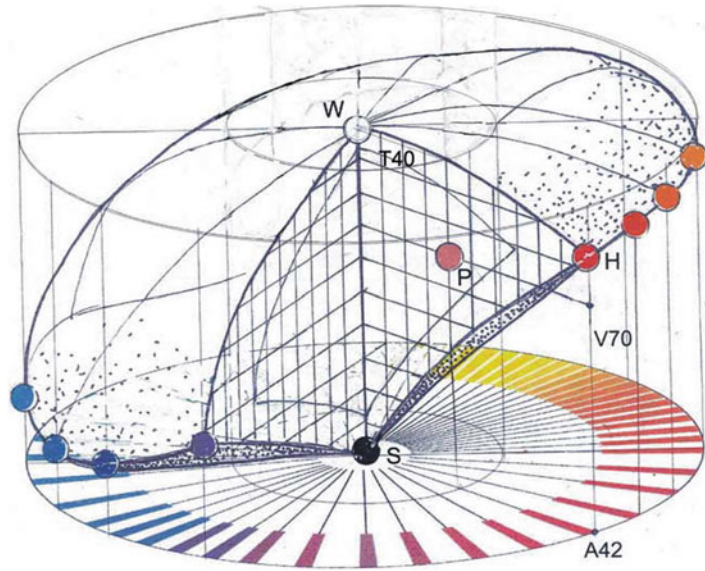
Color Order Systems, Fig. 10 Axial section of the coloroid solid, displaying one hue with the variation of lightness and saturation

yellow-cyan mixtures and 99% of magenta and the other with yellow-magenta mixtures and 99% of cyan.

Küppers also represented his system as a three-dimensional space, either as a cube (Fig. 12) or with a rhombohedral shape, constituted by an

Color Order Systems,

Fig. 11 Color solid of the coloroid system



upper tetrahedron, a central octahedron, and a lower tetrahedron (Fig. 13). In the rhombohedral model, white is placed in the upper vertex of the upper tetrahedron, in whose base the three subtractive printing primaries are placed: yellow, magenta, and cyan. This base is also one of the faces of the central octahedron. The lower triangular face of the octahedron holds the three colors resulting from the mixture of the subtractive primaries in pairs: green (yellow with cyan), red (yellow with magenta), and blue (cyan with magenta). This face is, in turn, shared with the lower tetrahedron, in whose lower vertex black is placed.

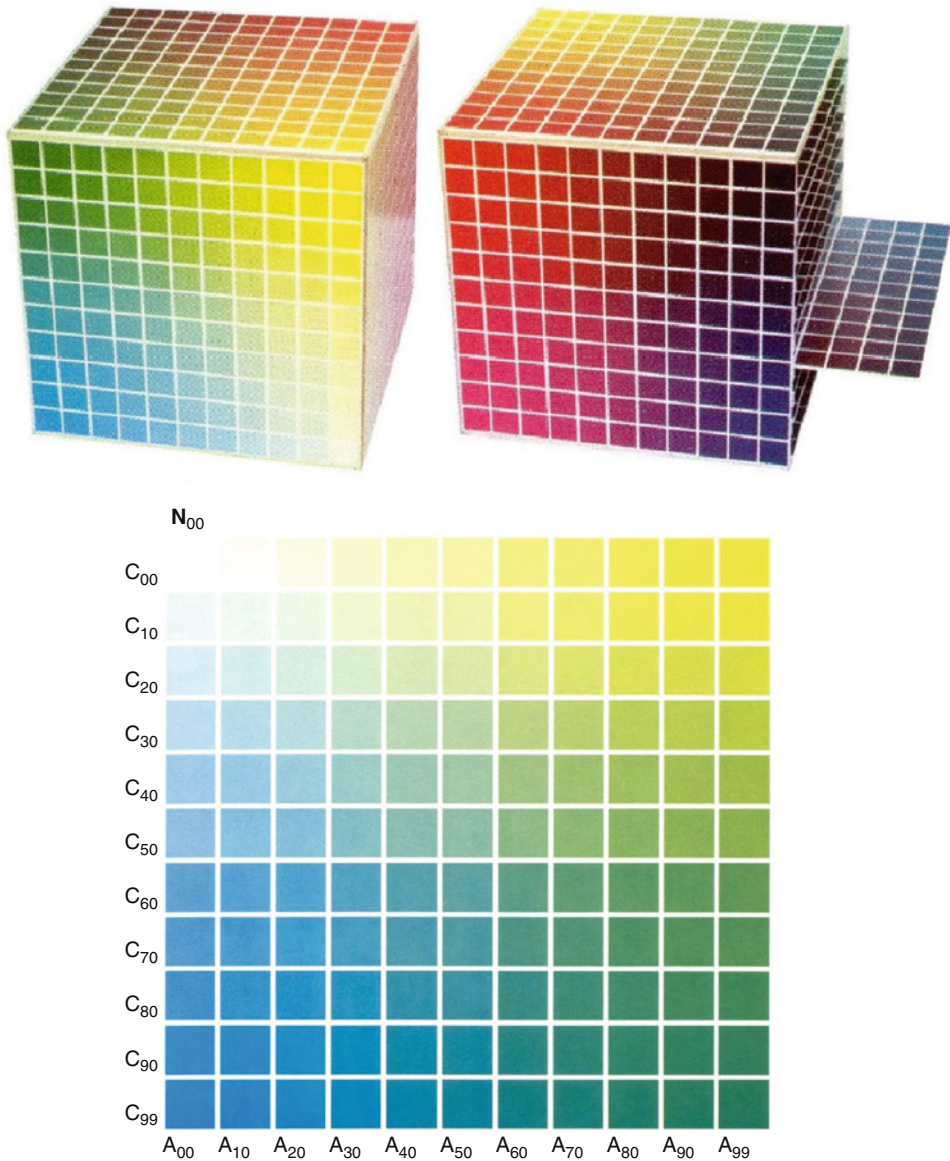
Additional Studies, Classifications, and Evaluations of Color Order Systems

The bibliography about historical, comparative, or classificatory analysis of color order systems is extensive. In addition to the already stated references, articles by Spillmann [13], Robertson [14], Sivik [15], and Tonnquist [16] offer significant information. The International Color Association devoted a whole congress to the theme of color order systems in 1983 [17] and had a study group on this subject from 1978 to 1990, chaired successively by ► Günter Wyszecki, ► Fred

Billmeyer, ► James Bartleson, and William N. Hale (see [18]).

Especially interesting is the analysis made by Billmeyer [19], which deals with the history and the principles of various systems, establishes comparisons and differences, and offers data about the attempts of conversion among the notations of various systems. William Hale [20] discussed the uses of color order systems, including those that have a relevant theoretical or scientific basis as well as those more pragmatic and with a specific commercial purpose. Tonnquist [21] offered another criterion to classify color order systems. He divided them into (a) *physical*, those that have actual samples as reference and that only exist in terms of an atlas or a color chart; (b) *psychophysical*, those whose definition is given by means of valences for a set of points in a color space, and whose definition is only valid to a combination of observer, illuminant, and measuring system, as, for instance, the CIE system; and (c) *perceptual*, those defined in terms of elementary color percepts, mental references of certain basic colors which serve to describe all the remaining colors, as, for instance, the NCS.

Some countries have adopted certain color systems as national standards. During the 6th Congress of the International Color Association, a specific discussion about the question if any



Color Order Systems, Fig. 12 Harald Küppers' model atlas in a cubic shape, and a page from it

particular system is better than the others took place ([22], vol. I, pp. 163–172). The conclusion was that there is no particular system that is best for all the applications, which may cover fields so different as color education, artistic practice in various disciplines, professional practice in diverse branches of design (architectural, graphic, industrial, textile, landscape design), specification of color in materials so dissimilar as paper,

fabrics, leather, plastics, metals or in complex industries such as the automotive, the food industry, color reproduction in TV, video, and computer displays; some systems are more useful than others for certain specific problems. In the 7th Congress of the AIC, a round table on color order systems was also organized ([23], vol. A, pp. 173–174) in which there was a consensus on some points, among them (a) that physical color



Color Order Systems, Fig. 13 Küppers' rhombohedric model

samples, although useful, are not essential to a color system and (b) that broadly, the systems can be divided into two types: systems of color appearance (a typical example would be the NCS) and systems of color stimuli (an example would be the CIE system). All systems developed so far are applicable to trichromatic observers. Monochromatic, dichromatic, and tetrachromatic observers have as yet no systems applicable to them.

The interested reader can look for additional information on description, analysis, and comparison of color order systems in Caivano [24], Stromer and Baumann [25], Silvestrini, Fischer and Stromer [26], Stromer [27], Kuehni [28], Kuehni and Schwarz [29], and Spillmann [30], among other sources. There is also an excellent website devoted to this subject: <http://www.colorsystem.com>.

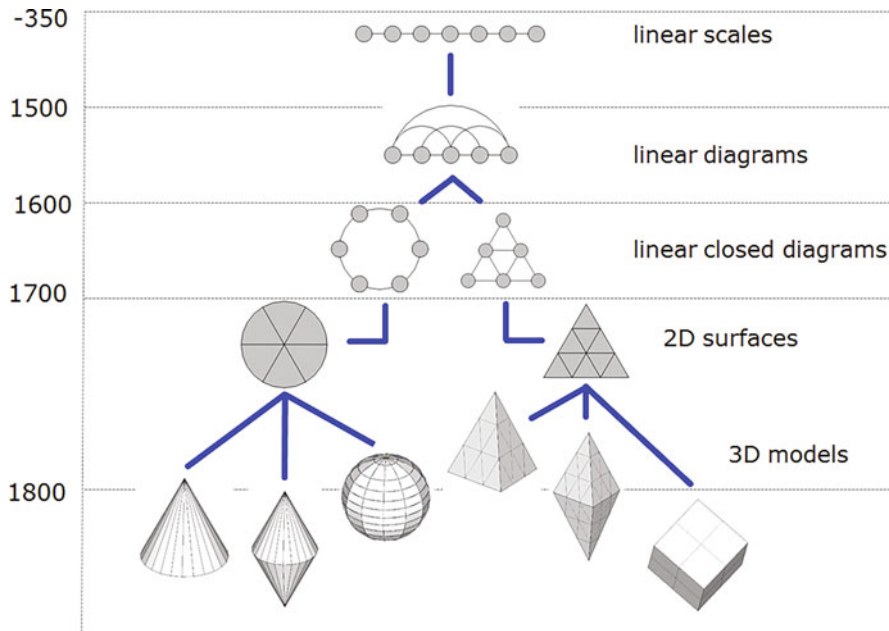
Punctuated Evolution of Color Order Systems

We can consider that the development of color order systems along history follows an

evolutionary path. However, instead of being a matter of gradual transformation, it is a kind of punctuated evolution, with steps or jumps where the changes appear.

With regard to spatial or geometrical aspects, linear arrangements (Aristotle and followers) are seen first. In the next step, connections are added to the linear sequence until it is closed in a circumference (Fludd) or a triangle (Zahn), but with empty spaces inside. Next, the circle was filled in (Newton), and achromatic colors disappeared from the border and were enclosed. This was a revolutionary step in the eighteenth century. While it took nearly 2000 years to close the circle, and near 100 years to fill it, it took just a little more than 50 years for this surface to be transformed into a three-dimensional space (Mayer, Lambert). These two last important changes took place within the eighteenth century, where the modern conception of color systems was developed (Fig. 14).

During the nineteenth and twentieth centuries, the 3D models suffered improvements: Additive and subtractive processes were separated, and the shapes started being divided into models with one vertex and models with two vertices.



Color Order Systems, Fig. 14 Scheme of the evolution of color order systems

Some of them began to appear with the chromatic circle tilted (pure yellow closer to white, pure violet, or purple closer to black), and thus, this surface became more elliptical than circular. Already in the twentieth century, some models, instead of having a convex volume (spheres, cones, double cones, pyramids, double pyramids), started to have more complex shapes, with positive and negative curvatures, both in the chromatic surface of full hues and in the outer surface of the model.

This process of punctuated transformation by means of more or less small steps or divisions will undoubtedly continue. What will be the next steps? We can think of at least two ways:

1) Adding one or two more dimensions to the 3D shapes, in order to encompass not only the three classical color variables (e.g., hue, lightness, saturation) but also other variables of **appearance**, such as gloss and transparency.

2) Taking into account that human color vision will probably evolve towards a tetrachromatic system (cases have already been found of women with **tetrachromatic vision**), and thus color order systems should reflect this tendency ... in a somehow distant future, however.

Cross-References

- ▶ [Al-Tusi, Nasir al-Din](#)
- ▶ [Appearance](#)
- ▶ [Avicenna, Abū ‘Alī al-Ḥusayn ibn ‘Abd Allāh ibn Sīnā](#)
- ▶ [Bartleson, C. James](#)
- ▶ [Billmeyer, Fred Wallace, Jr.](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIECAM02](#)
- ▶ [CIELAB](#)
- ▶ [Color Circle](#)
- ▶ [Goethe, Johann Wolfgang von](#)
- ▶ [Hering, Karl Ewald Konstantin](#)
- ▶ [Itten, Johannes](#)
- ▶ [Lambert, Johann Heinrich](#)
- ▶ [Mayer, Tobias](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Primary Colors](#)
- ▶ [Runge, Philipp Otto](#)
- ▶ [Tetrachromatic Vision](#)
- ▶ [Wyszecki, Gunter](#)

References

1. Shamey, R., Kuehni, R.G. (eds.): *Pioneers of Color Science*. Springer, Cham (2020)
2. Kirchner, E., Bagheri, M.: Color theory in medieval lapidaries: Nishaburi, Tusi and Kashani. *Centaurus*. **55**, 1–19 (2013)
3. Parkhurst, C., Feller, R.L.: Who invented the color wheel? *Color. Res. Appl.* **7**(3), 217–230 (1982)
4. Spillmann, W.: *Farbskalen – Farbkreise – Farbsysteme*. Applica, Sonderdruck, Wallisellen, Switzerland (2001)
5. Lang, H.: Trichromatic theories before young. *Color. Res. Appl.* **8**(4), 221–231 (1983)
6. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
7. Munsell, A.H.: *Munsell Book of Color*. Munsell Color Co., Baltimore (1942)
8. Hård, A., Sivik, L.: A theory of colors in combination: a descriptive model related to the NCS color order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
9. Nemcsics, A.: Coloroid colour system. *Color. Res. Appl.* **5**(2), 113–120 (1980)
10. Nemcsics, A.: Color space of the Coloroid color system. *Color. Res. Appl.* **12**(3), 135–146 (1987)
11. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York (1993)
12. Küppers, H.: *Das Grundgesetz der Farbenlehre*. DuMont, Cologne (1978) (English transl. *The Basic Law of Color Theory*. Barron's Educational Series, Woodbury, New York) (1982)
13. Spillmann, W.: Color order systems and architectural color design. *Color. Res. Appl.* **10**(1), 5–11 (1985)
14. Robertson, A.R.: Principles of colour order systems. In: AIC Colour 93, Proceedings of the 7th Congress, vol. A, pp. 149–153. Hungarian National Colour Committee, Budapest (1993)
15. Sivik, L.: Systems for descriptive colour notations – implications of definitions and methodology. In: AIC Colour 93, Proceedings of the 7th Congress, vol. A, pp. 89–94. Hungarian National Colour Committee, Budapest (1993)
16. Tonnquist, G.: 25 years of colour with the AIC – and 25000 without. *Color. Res. Appl.* **18**(5), 353–365 (1993)
17. AIC (Association Internationale de la Couleur): *The Forsius Symposium on Colour Order Systems*, 2 vols. Scandinavian Colour Institute, Stockholm, Colour Reports F26 and F28 (1983). Available at <https://aic-color.org/publications-proceedings>
18. Billmeyer Jr., F.W.: *AIC Annotated Bibliography on Color Order Systems*. Mimeoform Services, Beltsville (1987)
19. Billmeyer Jr., F.W.: Survey of color order systems. *Color. Res. Appl.* **12**(4), 173–186 (1987)
20. Hale, W.N.: Color order systems and color notations. In: AIC Color 89, Proceedings of the 6th Congress, vol. 1, pp. 43–51. Grupo Argentino del Color, Buenos Aires (1989)
21. Tonnquist, G.: Colour order systems and colour atlases. In: AIC Color 89, Proceedings of the 6th Congress, vol. 2, pp. 162–165. Grupo Argentino del Color, Buenos Aires (1989)
22. AIC (Association Internationale de la Couleur): AIC Color 89, Proceedings of the 6th Congress, 2 vols. Grupo Argentino del Color, Buenos Aires (1989). Available at <https://aic-color.org/publications-proceedings>
23. AIC (Association Internationale de la Couleur): AIC Colour 93, Proceedings of the 7th Congress, 3 vols. Hungarian National Colour Committee, Budapest (1993). Available at <https://aic-color.org/publications-proceedings>
24. Caivano, J.L.: *Sistemas de Orden del Color*. Facultad de Arquitectura, Diseño y Urbanismo, UBA, Buenos Aires (1995). Available at <https://colorsemiotica.files.wordpress.com/2015/11/1995scol.pdf>
25. Stromer, K., Baumann, U.: *Color Systems in Art and Science*. Regenbogen, Konstanz (1996)
26. Silvestrini, N., Fischer, E.P., Stromer, K.: *Farbsysteme in Kunst und Wissenschaft*. DuMont, Cologne (1998)
27. Stromer, K.: *Farbsysteme*. DuMont, Cologne (2002)
28. Kuehni, R.G.: *Color Space and its Divisions: Color Order from Antiquity to the Present*. Wiley, New York (2003)
29. Kuehni, R.G., Schwarz, A.: *Color Ordered: a Survey of Color Systems from Antiquity to the Present*. Oxford University Press, Cary (2008)
30. Spillmann, W. (ed.): *Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann*. Schwabe, Basel (2009)

Color Palette

► Color Scheme

Color Perception and Environmentally Based Impairments

Galina V. Paramei

Department of Psychology, Liverpool Hope University, Liverpool, UK

Synonyms

Acquired color vision impairment; Acquired color vision loss; Dyschromatopsia

Definition

Decreased discrimination of colors caused by adverse environment, such as long-term occupational exposure to or consumption of drugs, substances, and food containing neurotoxic chemicals.

Types of Neurotoxic Substances Causing Dyschromatopsias

Color vision early manifests adverse effects of exposure to an environment that contains neurotoxic substances [1–3]. The acquired color vision impairments, or dyschromatopsias, can be very subtle (subclinical) but also may vary considerably in severity, increasing or decreasing as long as the responsible agent persists, and can become irreversible under long-term exposure and/or agent dose.

There are several scenarios of exposure to hazardous chemical agents in the environment:

- (i) Long-term occupational exposure to certain substances (e.g., neurotoxic metals, organic solvents, carbon disulfide, etc.)
- (ii) Self-administered chronic consumption of substances containing neurotoxic chemicals (e.g., alcohol, tobacco).
- (iii) Side effects from pharmacological treatment of medical conditions (e.g., cardiovascular, antiepileptic, or antituberculosis drugs).
- (iv) Consumption of food contaminated through the food chain by neurotoxic elements (e.g., mercury).

General Characteristics of Neurotoxin-Induced Color Vision Impairments

Acquired color vision defects, unlike congenital ones, are noticeable to the observer: recently affected subjects name the stimuli as they see them – in contrast to subjects with congenital color vision defects where there is compensatory adaption of their color naming to that of normal trichromats. Acquired

dyschromatopsia may not be identical in the two eyes, which requires testing the two eyes separately.

Neurotoxic substances can affect one or more loci in the color vision system. At the pre-receptor level, hazardous chemicals can accelerate yellowing of the crystalline lens which results in an increase in absorption of short-wavelength light and hence decreased discrimination of blue colors. In the retina, the main mechanism of color vision loss is selective damage to specific photoreceptor classes, short-wavelength (S-), middle-wavelength (M-), or long-wavelength (L-) cones (► [Cone fundamentals](#)). Most vulnerable among these are the S-cones, damage of which is manifested by blue color vision defects [3]. Post-receptoral processing can also be disrupted – at the level of ganglion cells, optic nerve, optic radiation, or visual cortex – causing color vision impairment. Often the damage is nonselective, i.e., patterns of color discrimination loss are not always specific to one of the color subsystems and differ from those in congenital abnormalities.

Classification of Acquired Dyschromatopsias

The wide variation in acquired color vision defects, according to Verriest [4], can be classified in four major types, I, II, III, and a nonspecific defect. The first two types are associated with impaired color discrimination along the red-green axis in perceptual color space (Color Vision, Opponent Theory), much like the patterns found in congenital red-green deficiency, i.e., both involve mild to severe confusion of reds and greens. Type I is protan-like and reveals little or no loss of blue-yellow discrimination (► [Protanopia](#)); Type II is deutan-like and is manifested by concomitant mild loss of discrimination between blues and yellows (► [Deutanopia](#)). Type III, tritan-like, is manifested by mild to moderate blue-green and yellow-violet confusions, with a lesser or absent loss of red-green discrimination (► [Tritanopia](#)). Both color systems appear to be selectively susceptible to damage by various types of neurotoxins.

According to Köllner's rule [1], impairment of blue-yellow discrimination – the range most frequently affected by exposure to hazardous chemicals – suggests toxic retinopathy, i.e., a more external retinal dysfunction; by contrast, a preponderance of red-green loss is associated with pathology in the optic nerve; finally, complex color vision loss, blue-yellow and red-green, suggests a more advanced stage with a damage to both the retina and the neuro-optic pathway. However, numerous exceptions to Köllner's rule instruct one to be cautious about making a clear-cut attribution of blue-yellow loss to damage at a retinal level and of red-green loss to damage at a neural level.

Tests of Color Vision for Assessing Acquired Dyschromatopsias

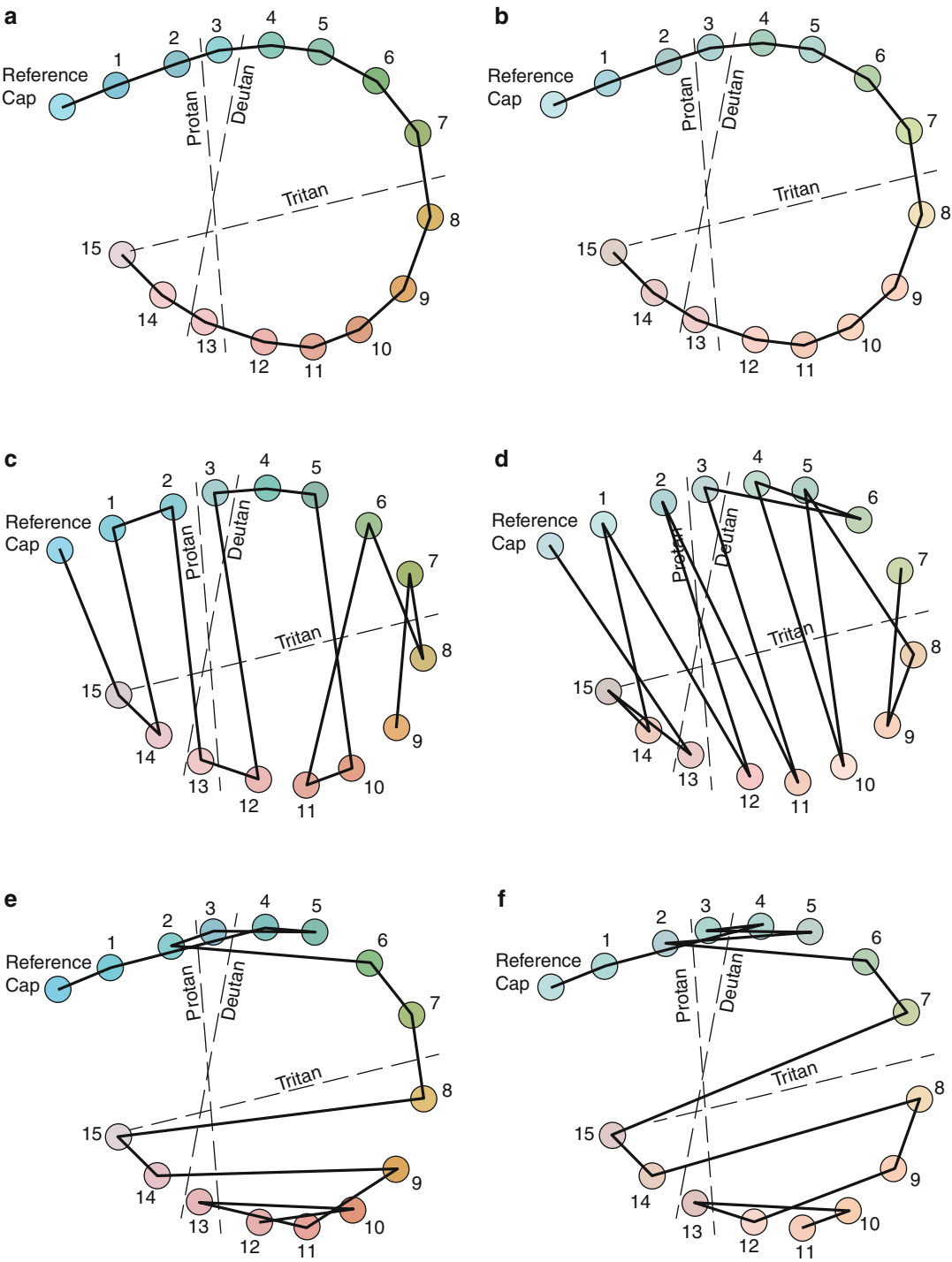
In epidemiological studies, three color arrangement tests are predominantly used: the Farnsworth-Munsell 100 Hue (FM100) test, the Farnsworth Dichotomous panel D-15 test, and the Lanthony Desaturated Panel D-15d (► [Color Vision Testing](#)). In the arrangement tests, caps sample a color circle (► [Color Circle](#)) at even intervals. In field studies, the two shorter tests, the D-15 and the D-15d, are predominantly employed that can be rapidly administered. The two tests are often used in conjunction, while the more sensitive D-15d is widely employed for early detection of mild neurotoxin-induced dyschromatopsias. The observer is presented with a set of color caps and requested to arrange them in ("rainbow") sequence. The D-15 test was designed to diagnose moderate to severe color defects; the Lanthony D-15d test, similar in design and identical to the D-15 test in administration, consists of color samples that are lighter and paler [6] and was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. Outcomes of both tests are easily interpreted allowing color vision ability to be quantified graphically [5]. Further, the number of erroneous cap transpositions provides a measure of overall color discrimination; the pattern of the transpositions indicates whether the defect is

closer to the blue-yellow or red-green axes or with no discernible pattern. The outcomes of the D-15 and D-15d are reported via a Color Confusion Index (CCI), where 1.0 corresponds to perfect color arrangement; CCI values greater than 1.0 indicate progressive impairment of color discrimination [5, 6] (Fig. 1).

Occupational Exposure to Neurotoxic Substances

A number of occupations involve exposure to volatile neurotoxic substances (e.g., printers, aircraft maintenance workers, dry cleaners in automotive and metalworking industries, viscose rayon workers, microelectronics assembly workers, gold miners, dentists, etc.). Such substances include organic solvents (toluene, styrene, benzene, perchlorethylene, *n*-hexane, carbon disulfide), solvent mixtures, and metals like mercury (in its elemental or methyl forms). Even when neurotoxic substances are applied within the occupational limits, long-term exposure has been shown to result in mild impairment of color discrimination [7–9].

Using the D-15d test, the degree and pattern of color vision loss were intensively investigated with regard to exposure to organic solvents, in particular toluene and styrene [2, 7–10], and to mercury [11]. The main finding across these studies is significant increase of the CCI in the occupationally exposed observers (compared to age-matched controls). For example, in a meta-analysis of 15 sample studies of the effects of toluene, styrene, and solvent mixtures [9], the grand mean CCI for the exposed workers was 1.22 ± 0.08 , significantly greater than 1.13 ± 0.06 for the controls ($p = 0.003$). Similarly, in [11], for workers of fluorescent lamp production exposed to mercury vapor, $CCI = 1.14 \pm 0.14$ was significantly greater than $CCI = 1.04 \pm 0.06$ for controls ($p = 0.002$). The impairment of color discrimination was shown to be subject to cumulative exposure, i.e., product of duration and current level of exposure [2, 7–10], and may become irreversible even when the hazardous agent is withdrawn [11].



Color Perception and Environmentally Based Impairments, Fig. 1. Scoring sheet for the D-15 (a) and the D-15d (b); for illustrative purposes, the numbers are accompanied by colors simulating those of the original test caps. At perfect color arrangement of a normal trichromat, lines connect the “reference cap” through 1–15; CCI = 1.0. (c, d) Lines are drawn between consecutive

caps as placed by a protanope, observer with a congenital red-green deficiency; (c) CCI = 2.48; (d) CCI = 3.11. (e, f) Mild acquired tritan (blue-yellow) type of color discrimination impairment; (e) CCI = 2.06; (f) CCI = 2.19. Note that the low-saturated stimuli of the D-15d result in a more prominent color confusion (d, f)

Neurotoxin-induced dyschromatopsias are predominantly of Type III, i.e., of tritan-like pattern. Less common are Type I and II, red-green dyschromatopsias. In comparison, the non-specific type of dyschromatopsia is also highly prevalent, which implies difficulty in discriminating colors along both the red-green and blue-yellow axes of color space (Color Vision, Opponent Theory).

Self-Administered Consumption of Substances Containing Neurotoxic Chemicals

Chronic excessive consumption of alcohol (ethanol) affects color discrimination capacity [6, 7]. When assessed by the D-15d test, the prevalence of dyschromatopsia was shown to increase with alcohol intake. Further, heavy drinkers (with an intake larger than 750 g/week) manifested primarily loss of blue-yellow discrimination, whether or not they were undergoing treatment in a detoxification center. However, 25% of persons undergoing detoxification revealed dyschromatopsia of the nonspecific type, including red-green loss [12].

Tobacco smoke contains a range of compounds including nicotine, cyanide, and carbon monoxide and, when consumed excessively, can affect color vision [1, 13]. When tested by means of the color arrangement tests, chronic smokers (>20 cigarettes/day; for at least a year) revealed a subtle but statistically significant reduction in sensitivity to red-green differences compared to non-smokers [14] or showed a diffuse character of color vision disturbance, without a particular dyschromatopsia axis [15].

Side Effects from Pharmacological Treatments

A number of medications (e.g., cardiovascular, antiepileptic, antituberculosis, antirheumatism drugs, oral contraceptives, etc.) are known to

produce measurable color vision disturbances, some of which affect color vision even at therapeutic levels, most commonly as Type III, tritan-like blue-yellow dyschromatopsia [1, 3, 6, 7, 16]. For instance, patients taking antiepileptic drugs develop mild blue-yellow deficiency which may show signs of progression with lasting intake of the drugs. Intake of Viagra was shown to cause transient adverse visual events described as a blue color tinge to vision, accompanied by mild blue-yellow deficiency in about 11–14% of those taking the medication, the disturbance being reversible after the medication has been discontinued. Treatment by the tuberculostatic ethambutol shows mild blue-yellow deficiency as the earliest sign of the drug's neurotoxicity, but this can also develop as Type II, deutan-like red-green, or a nonspecific dyschromatopsia; these defects are transient and reversible after stopping the therapy.

Consumption of Food Contaminated by Neurotoxic Elements

Certain industrial activities, like gold mining or mercury mining, are associated with pollution of mercury which bioaccumulates mainly through the aquatic food chain (seafood and fish). Even at low levels of dietary exposure, mercury was shown to chronically reduce color discrimination, with the error pattern indicating that both blue-yellow and red-green systems are affected, when the FM100 test, D-15d test, and the Cambridge Colour Test were used (► [Color Vision Testing](#)) [17, 18].

Cross-References

- [Color Circle](#)
- [Color Vision Testing](#)
- [Cone Fundamentals](#)
- [Deuteranopia](#)
- [Protanopia](#)
- [Tritanopia](#)

References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L.G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Gobba, F.: Color vision: a sensitive indicator of exposure to neurotoxins. *NeuroToxicology* **21**, 857–852 (2000)
3. Simunovic, M.P.: Acquired color vision deficiency. *Surv. Ophthalmol.* **61**, 132–155 (2016)
4. François, J., Verriest, G.: On acquired deficiency of colour vision, with special reference to its detection and classification by means of the tests of Farnsworth. *Vis. Res.* **1**, 201–219 (1961)
5. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
6. Geller, A.M., Hudnell, H.K.: Critical issues in the use and analysis of the Lanthony Desaturate Color Vision test. *Neurotoxicol. Teratol.* **19**, 455–465 (1997)
7. Iregren, A., Andersson, M., Nylén, P.: Color vision and occupational chemical exposures: I. An overview of tests and effects. *NeuroToxicology* **23**, 719–733 (2002)
8. Gobba, F., Cavalleri, A.: Color vision impairment in workers exposed to neurotoxic chemicals. *NeuroToxicology* **24**, 693–702 (2003)
9. Paramei, G.V., Meyer-Baron, M., Seeber, A.: Impairments of colour vision induced by organic solvents: a meta-analysis study. *NeuroToxicology* **25**, 803–816 (2004)
10. Benignus, V.A., Geller, A.M., Boyes, W.K., Bushnell, P.J.: Human neurobehavioral effects of long-term exposure to styrene: a meta-analysis. *Environ. Health Perspect.* **113**, 532–538 (2005)
11. Feitosa-Santana, C., Costa, M.F., Lago, M., Ventura, D.F.: Long-term loss of color vision after exposure to mercury vapor. *Braz. J. Med. Biol. Res.* **40**, 409–414 (2007)
12. Mergler, D., Blain, L., Lemaire, J., Lalande, F.: Colour vision impairment and alcohol consumption. *Neurotoxicol. Teratol.* **10**, 255–260 (1988)
13. Fernandes, T.P., Silverstein, S.M., de Almeida, N.L., Santos, N.A.: Visual impairments in tobacco use disorder. *Psychiatry Res.* **271**, 60–67 (2019)
14. Bimler, D., Kirkland, J.: Multidimensional scaling of D15 caps: color-vision defects among tobacco smokers? *Vis. Neurosci.* **21**, 445–448 (2004)
15. Erb, C., Nicaeus, T., Adler, M., Isensee, J., Zrenner, E., Thiel, H.-J.: Colour vision disturbances in chronic smokers. *Graefe's Arch. Clin. Exp. Ophthalmol.* **37**, 377–380 (1999)
16. Zrenner, E., Hart, W.: Drug-induced and toxic disorders in neuro-ophthalmology. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.) *Clinical Neuro-Ophthalmology: A Practical Guide*, Ch. 17, pp. 223–229. Springer, Berlin/Heidelberg (2007)
17. Silveira, L.C.L., Damin, E.T.B., Pinheiro, M.C.C., Rodrigues, A.R., Moura, A.L.A., Cortes, I.M.I.T., Mello, G.A.: Visual dysfunction following mercury exposure by breathing mercury vapour or by eating mercury-contaminated food. In: Mollon, J.D., Pokorny, J., Knoblauch, K. (eds.) *Normal & Defective Colour Vision*, pp. 409–417. Oxford University Press, Oxford (2003)
18. Feitosa-Santana, C., Souza, G.S., Sirius, E.V.P., Rodrigues, A.R., Cortes, M.I.T., Silveira, L.C.L., Ventura, D.F.: Color vision impairment with low-level methylmercury exposure of an Amazonian population Brazil. *NeuroToxicology* **66**, 179–184 (2018)

Color Phenomenology

Don Dedrick

Department of Philosophy and Department of Psychology, University of Guelph, Guelph, ON, Canada

Synonyms

Color experience; Color qualia

Definition

The word “phenomenology” finds its original application in philosophy, and it has two distinct meanings in that discipline. In the first, most substantive meaning it refers to a philosophical tradition originating in the work of G. W. F. Hegel and developed in the work of Edmund Husserl, Martin Heidegger, Jean-Paul Sartre, and others, with the psychologist Franz Brentano as a major influence [11]. In this primordial sense, it refers to a non-psychological description of the fundamental constituents of experience. It may sound peculiar to call an account of experience “nonpsychological” since experience might be thought of as necessarily psychological. The rationale for the usage is this: one may possibly describe the fundamental constituents of human experience – concepts, ideas, propositions, temporality, mental images, etc. – in a way that captures their generic character and hence their “universality” rather than their specific contents. Such a view articulates, so classical phenomenologists say, the logical or

conceptual structure of experience. The second application of the term refers more directly to the *mere* appearance of things. There is no universality attached to such description. It is, rather, a reference to the way things *seem* to perceivers: red looks this way (and perhaps just to an individual), pain feels this way (ditto), dogs bark in a way that sounds as it sounds (again, perhaps just to the individual). This sense of the word “phenomenology” (the words “qualia” [10] or “experience” are sometimes used) describes the way that many Anglo-American philosophers have deployed the term throughout much of the twentieth century, into the twenty-first. This second sense has clear links to the way experimental psychologists talk about colour experience, especially when they refer to the psychological, sensory aspects of color: hue, saturation, lightness/brightness, and other aspects of color appearance [8]. It is this second sense that is of interest here.

Color Phenomenology and Ontology

If the phenomenology of X concerns the way that some X *seems*, is a contrast to the way things *are* implied? The answer is yes. Consider the following example (discussion will now concern examples and issues specific to color phenomenology). As one moves a color stimulus from one illumination source to another, one will notice that it appears to change color, though likely within the bounds of color constancy. This may lead to a question as to its “real color,” and that question may lead in a number of directions. Perhaps its real (object) color is identical with a physical property that does not change across illuminations: surface spectral reflectance, for example. Perhaps the idea of real color should be abandoned and replaced with a conception of color relativized to viewing conditions and observers. Perhaps one could speak of a “normal observer” (as the CIE does) so as to wring some objectivity out of the phenomenology.

These positions take up different views as to the relation between the way things seem – color phenomenology – and the way things are, ontologically, for color. There is a spirited, scientifically informed philosophical literature that covers

many permutations of three general views which are typically called, with some variation, *objectivism* (color is in the world), *subjectivism* (color is in the mind-brain), and *relationalism* (color is constituted in a relation between the mind-brain and the world). There is currently no consensus among philosophers as to which type of view is correct. All three exhibit strengths as well as weaknesses. To consider one example: an objectivist view allows us to distinguish between the colour of an object and its colour as perceived by a subject. That is an obvious strength. On the other hand, objectivist views find it difficult to specify an appropriate candidate for objective color. It is not just that there are many actual physical causes of color experience, but that the actual causes are often identified by reference to experience itself. The colour red, for example, seems to be, objectively, whatever it is that causes a red color experience. That is a weakness. (See [1, 5] for discussion of this and related issues.)

Epistemological Issues

Questions about real colors are questions about ontology. Does the catalogue of the real include colors? What is a real color, if there is such? As important as these questions are, much of the interest when it comes to color phenomenology is not ontological but epistemological: what can be known, and what are the limitations to what can be known, about color experience? The most famous query along this line comes from John Locke. In the *Enquiry Concerning Human Understanding* [6], Locke proposes that a “spectral inversion” would not be detectable. Subject A and subject B, indistinguishable in terms of their behavior on discrimination tasks, nonetheless have different experiences. Subject A’s experiential color space is “inverted” relative to B’s. For Locke, this meant that A and B use color words the same way, describe their color experiences the same way, discriminate color stimuli in the same way, and yet have distinct color experiences, e.g., Subject A’s green is B’s red, and vice versa, and the same goes for blue and yellow, etc. Thus, the inversion is behaviorally undetectable. Many consequences have been

thought to flow from Locke's proposal (that, for instance, behaviorism is false as a theory of psychology), but vision science provides good reason to believe that even if such inversion were possible, it would be detectable. The inverted spectrum proposal depends on a color space that is symmetrical so that one can map discriminable differences one-to-one from, say, the "greens" to the "reds." This condition is not satisfied for a standard human trichromatic color space, as specified, for instance, in the asymmetrical CIE $L^*a^*b^*$ space. The upshot, in terms of behavior, is that A and B would behave differently – confusing or discriminating different color stimuli. The inversion, with the cleverness of psychophysics, would be detected [2].

Despite the fact that Locke's proposal appears to fail, its implications are not easily dispatched. Even if spectral inversion can be detected, the question remains as to what color experience is like. Consider the following line of reasoning. What is to be learned from the psychophysics of color? This is a broad question, but broadly the answer would be that one learns correlations between different types of stimuli and different types of behavior. From these correlations, serving as constraints, ideas as to what the properties of the neural substrate of color experience need to be like may be formulated, models constructed, convergence with other areas of physiology and psychology sought. Psychophysics (visual being the concern here) is a mature subdiscipline of psychology, but does it deal with the way things – colored things – look? This sounds an odd question for a science that is based on subjects looking at visual stimuli and responding to them. How could it not deal with this? Yet vision scientists are uneasy about the claim that discrimination data say something reliable about the content of experience. If one asks a subject what red looks like, the subject will revert to demonstrative claims – it looks like *that* – or to relational claims locating a red color sample in relation to that of other colors: more like orange and yellow, less like blue and green. While these descriptors are often robust for subjects, they do not, or so it is often claimed, get at the subjective nature of color experience. In a very influential article, "What is it like to be a bat?" [7] Thomas Nagel argued that we

cannot access the subjective experience of others from a scientific third-person perspective. Bat's, Nagel proposed, have experience – there is something that it is like to be a bat (as opposed to, say, a stone). While one can use the techniques of animal ethology, biology, and bat psychophysics to determine bat discrimination space, the content of the bat's experience is beyond the grasp of those third-person methods – beyond an objective third person science and its "view from nowhere." Nagel's argument really has little to do with bats. They are a useful exemplar since it is easy to imagine (a) that bats have experience and (b) that their experience is distant from human experience. In this sense, the bat is a foil to arguments from analogy: one can believe that bats have experience, but there is no analogy to human experience to guide us because echolocation is a different sensory modality than either sight or hearing. The central thrust of Nagel's argument concerns subjectivity. There is a subjective view of the world that science, as humans presently know it, cannot access. Human subjectivity is, in the end, just as mysterious from the third person point of view as is the bat's. This is why many psychophysicists are likely to be in agreement with philosophers skeptical about our knowledge of color phenomenology – why it is that psychophysicists are uneasy about any inference from discrimination to experience – inference from the judgments subjects make on discrimination tasks to the nature of their experience.

Not all philosophers, and certainly not all scientists, vision or otherwise, accept the view that subjective experience – including color experience – is mysterious from a third-person perspective. One might argue that a good model of an individual's discrimination space for color is as good as the human behavioral sciences get. If, for example, a subject fails to discriminate images in some set of pseudoisochromatic color plates, then one can make predictions about their future discriminatory behavior and also explain that behavior. Such tests do more than identify types of "color blindness"; they identify the axes on which colors are confused and may correlate such confusion with genetic differences in opsin expression at the retinal level. What more could

one want? It seems that there are two things: (1) an account of *what* the subjective experience of color is like, as opposed to a third-person take on that experience – this is Nagel’s concern and (2) an account of *how* subjective experience is generated (how it fits in with the ontological “catalogue of the real”) – this is a concern most closely identified with the Australian philosopher David Chalmers [3].

Chalmers writes mainly about consciousness, but his views on that subject have clear implications for color phenomenology. Unlike Nagel, who makes a case for the subjectivity of experience and is concerned with how that experience might be understood scientifically, Chalmers argues that the real problem with experience is that science has no idea as to how it is generated by a physical system (and, more radically, *why* there should be such experience in the first place). Chalmers is not denying there is subjective experience. We have experiences of color, unquestionably. He is claiming that the causal account we give of such experience is incomplete. Suppose one could understand color perception “all the way up”: from stimulus presentation, to photon-absorption at the retinal level, to retinal and LGN opponent processing, to cortical processing in the visual areas of the brain, to integrative processing in the executive areas, to the output of discrimination-based behavior which is a function of this whole process. While vision science understands some elements of this causal story quite well, and others not as well, even *perfect* understanding of it might leave science in the dark as to how color experience is generated. At what point do the biological properties of brains cause or constitute experiences of color, and how? Chalmers argues that the science of consciousness, such as it is, has no idea how to even address, let alone answer this so-called “hard question” – “hard” not in virtue of the difficulty of the science (as with the molecular biology of vision, say) but “hard” in the sense that science offers *no advice* on how to bridge the gap between its cognitive-neurobiological accounts of the brain, and the experience that brain is thought to generate. In response to the idea that there is such a hard problem, critics of Chalmers have typically

adopted one of two approaches (see the commentaries reprinted in Ref. [9]): (1) argue that Chalmers assumes the limitations of current brain science are permanent (this involves the positive claim that problems that look “hard” at a given point in time may become easier with new developments in theory and practice) or (2) argue that early progress has, even now, been made on a complete theory of conscious experience.

With respect to the first strategy, one can quite agree that, in the future, science is likely to succeed in areas currently unimagined. Yet (1) does not address the request for a way forward on the problem of brains generating experience but merely points out that science will almost certainly find new ways of addressing (or disposing of) the problem. This may be true, but the argument is not substantive, given the claim that the nature of experience and its origins is a different sort of problem altogether. If one claims that the march of science will be sufficient to explain the generation of conscious experience at some point in time, then one is simply denying Chalmers’ claim and that, arguably, assumes what it should demonstrate. As for (2), the view that progress has been made on the explanation of consciousness, Chalmers has pointed out that contemporary empirical theories of conscious experience (a) shift the problem of experience and how it is caused or constituted to accounts wholly within cognitive neuroscience and biology – an account of attention, say, or an account of neural opponent processing for the case of color. As a consequence (b), such accounts will provide, at best, accounts of the “neural correlates” of conscious experience rather than an account of how conscious experience is generated by neurobiological function.

Color Phenomenology and the Structure of Color Experience

Claims that science does not understand how phenomenological experience arises or is constituted from neurobiological function, or that a complete understanding of human phenomenology surpasses what can be known from a third-person scientific perspective sound dire: as if experience

is not understood at all. And yet it is remarkable how much about the phenomenology of color can be examined “from the outside” – and from a position of ignorance as to the ultimate causes of that phenomenology. As noted earlier, psychophysics would be able to identify the spectral inversion that has troubled philosophers. Such detection really is no different, in principle, than the detection of different forms of dichromacy: protanopia, deuteranopia, and tritanopia – all involving failure to discriminate among stimuli that a normal trichromat would discriminate. Each of these deficiencies (relative to trichromacy) is a consequence of the lack of one or another of the three typical photoreceptors. They are physiologically based effects that can be identified through behavioral tests, and, moreover, vision science can explain the relevant deficiencies and their fine-grained differences at physiological and genetic levels. A color space for the dichromat, as for the trichromat, can be pieced together from various behavioral judgements, a difficult and arduous task. Color science has, in other words, a reasonably good grasp of the *structure* of human (primate) color experience [4], as well as its physical and neurophysiological correlates, even if the precise, personal, subjective nature of that experience as well as its generation by physical systems such as brains remain epistemically problematic.

Cross-References

- Appearance
- Color Order Systems
- Color Psychology
- Philosophy of Color
- Pseudoisochromatic Plates
- Psychological Color Space and Color Terms

References

1. Brown, D., Macpherson, F. (eds.): The Routledge Handbook of Philosophy of Colour. Routledge, Abingdon-on-Thames, UK (2019)

2. Byrne, A.: Inverted Qualia. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (Winter 2018 Edition) (2018). <https://plato.stanford.edu/archives/win2018/entries/qualia-inverted/>
3. Chalmers, D.: Explaining Consciousness: The Hard Problem. MIT Press, Cambridge, MA (2000)
4. Clark, A.: Sensory Qualities. Oxford University Press, Oxford (1993)
5. Cohen, J., Matthen, M. (eds.): Color Ontology and Color Science. MIT Press, Cambridge, MA (2000)
6. Locke, J.: In: Nidditch, P.H. (ed.) An Essay Concerning Human Understanding. Oxford University Press, Oxford (1689/1975)
7. Nagel, T.: What is it like to be a bat? *Philos. Rev.* **83**, 435–450 (1974)
8. Oleari, C.: Perceptual phenomenology of light and of color (Ch. 4). In: Standard Colorimetry: Definitions, Algorithms and Software. Wiley, Chichester (2015)
9. Shear, J. (ed.): Explaining Consciousness: The Hard Problem. Bradford Books/MIT Press, Cambridge, MA (1999)
10. Tye, M.: Qualia. In: Zalta E.N. (ed.) The Stanford Encyclopedia of Philosophy (Summer 2018 Edition) (2018). <https://plato.stanford.edu/archives/sum2018/entries/qualia/>
11. Zavhi, D.: The Oxford Handbook of the History of Phenomenology. Oxford University Press, Oxford (2018)

Color Pollution

Malvina Arrarte-Grau
Architecture, Landscape and Color Design,
Arquitectura Paisajismo Color, Lima, Peru

Synonyms

Inappropriate use of color; Visual contamination;
Visual pollution

Definition

Color pollution consists of an inappropriate color arrangement which causes or increases disorder in the perception of the visual field within an urban or natural environment. It is an important aspect of visual pollution.

Overview

Color Pollution as a Component of Visual Pollution

In the context of environmental design, visual pollution refers to all nonarchitectural elements which spoil, in an invasive and simultaneous way, the perception of outdoor spaces. These features range from plastic bags trapped between tree branches to publicity panels, street signs, posts, and wires [1]. Besides the unaesthetic consequences the unnecessary exposure of these elements may bring, visual pollution leads to the overstimulation of the senses, increasing the load of information to be processed by viewers, drivers, and pedestrians.

Light and color are important aspects of visual pollution. The first one generates luminous pollution or light pollution due to artificial lighting and the second one generates color pollution by the introduction of cultured elements. Color pollution may originate in visual pollution, when it is the consequence of an inconveniently positioned feature which reinforces its presence by color. Otherwise, it may be a color that is incoherent within the composition and causes disorder. The disharmonious result may be produced by one or more color dimensions (hue, saturation, and lightness) [2]. Achromatic colors are necessary for creating transition spaces between color information, though black, white, and gray may also produce an effect of color pollution by lack of tone and contrast in luminosity.

Negative Impact of Color in the Environment

Color affects the perception of objects and compositions within a setting. It serves to codify elements and establish visual hierarchies. If used randomly, disregarding its power, color may easily draw attention to misleading data, distracting the viewer from relevant information for the interpretation of a scene [3]. When inappropriate colors are used in minor elements such as urban furniture, advertising panels, or building elements, color, far from being a useful signal, ends up invading the environment with an ambiguous and arbitrary presence.

Color as Intended by Nature and Adaptation to Color Coding

Colors in nature convey physical and chemical characteristics. They also give information on material processes and ethereal substances. In the animal kingdom, color has a biological role, as an aid to survival [4]. In natural conditions, human beings would respond to colors by instinct, using them as intended by nature: for alerting and announcing danger, for recognizing food and bodily functions, and for perceiving space and distance.

The man-made environment is full of visual information and signals, many of which are based on color. Universal coding by color distinguishes hot from cold, gas from water, forbidden from allowed, and so forth. Color is present in domestic and work activities. When shopping, colors in packaging and logos inform about names and brands, colored tags serve to mark sales from regular prices, while bright and fluorescent colors indicate special deals.

When color is used inadequately, the information it conveys becomes irrelevant and negative for the user, causing confusion and contributing to visual pollution. Man possesses a biological memory that keeps him aware of color signals, but using this faculty amid color excess and randomness may prove a waste of energy, causing fatigue and stress.

Color Pollution in Cities

Color pollution exists in the urban environment, predominantly in cities and villages. Commercial and mixed-use areas are good examples of competition for capturing attention. In isolation, color communication is effective. The use of written signs and subtle colors may work too, but among competing signs, these no longer accomplish their intended function, promoting other formulas to be put into practice: brighter colors, contrasting backgrounds, larger letter types, images, lighting, and many others [5]. The arrangement of the signs may easily become visual pollution, while the excessive and inappropriate use of color contributes to a chaotic environment.

The effects of color pollution are negative. Bright and saturated colors tend to spread in the

scene, as others are induced to use the same strategy. In cities, color tends to go out of control in commercial panels. Traffic and safety signs [2] diminish their effectiveness when there are objects with similar colors around. Associations between certain colors and objects, which have not been planned carefully, tend to become ingrained in a cityscape, that is, urban furniture, bridges, and signals. From that base, the rest is decided or added on.

Architectural elements may cause color and visual pollution within a façade by producing imbalance in the composition. When building exteriors invade the visual field with bright and saturated colors, far from enhancing architecture, these threaten the aesthetic aspect of the environment.

Color Pollution in Natural Settings

Natural settings are highly susceptible to visual and color pollution. This occurs when man-made elements are introduced in a landscape, with no regard to material quality and color, resulting alien and obtrusive. It is common to find posts, cables, advertising, and constructions invading the visual field, interrupting the continuity of mountain ranges, seascapes, agricultural fields, and woods. The position, size, and frequency of these elements are relevant. Additionally, their impact may be emphasized or diminished by materials and colors. For instance, a black plastic water tank may be very disruptive when placed against a mountain backdrop. Its visual impact would be attenuated if it were beige, echoing the color of the background.

Consequences of Color Pollution

As occurs with noise and visual contamination, color pollution may be aggressive, causing fatigue and stress. Slow reactions and traffic accidents may be caused by excess of information and distracting colored elements. As color pollution affects the aesthetic aspect of a place, users may find it hard to enjoy and develop a sense of belonging in visually disrupted settings.

Though the perception of color in animal species differs from that of humans, pollution by

color may diminish the chances of wildlife establishment near human conglomerates. In prey birds and day birds, which have an acute and sophisticated sense of vision, color plays specific functions in mating and feeding [2, 6]. The presence of artificially colored elements in their habitat may affect their behavior by alerting and confusing them.

Causes of Color Pollution

If color does not correspond to function nor to survival or aesthetics, who is to blame for its randomness? Lack of regulation and guidelines for the use of color allow the invasion of contaminating colored elements in the environment. This is a major deficiency of cities and natural settings. The causes of color pollution also rely on the consumer society, the commercial offer of paint and cladding materials, and visual marketing strategies.

Man, with his ability to synthesize and create from the materials which are granted to him, uses color to express, symbolize, and communicate ideas. The evolution of coloring mediums, tools, and technologies has resulted in an overwhelming variety. By contrast, the rules for the application of these advancements are scarce.

Municipalities rent strategic spaces to corporations for publicity. Owners will rent the roofs of their houses to advertising companies, if they are not prevented to do so. To the detriment of the setting, a gigantic beer can or a credit card could become a focal point. The colors used in these advertisements are centered in the product, not in the environment, and very often result in color pollution.

It appears as common practice in many parts of the world that, at the time of political elections, propaganda invades city and countryside. Usually the combination of two primary colors of medium to high saturation and black is used over a white background, producing a contrasting figure-ground relationship which is legible from a distance [5]. Recurring examples of visual and color pollution of this sort remain on walls and along roadsides long after the campaigns are over.

It is for sure that the cultivated and sensitive individual values unspoiled scenery more than

authorities and inhabitants of rural towns, where the beauty of natural sights seems to be taken for granted.

Solutions for Color Pollution

Authorities have in their hands the possibility to control the environmental impact caused by political propaganda and commercial advertising through regulations and sanctions.

Decisions for preserving and improving environmental colors are common ground of authority and designer. It is necessary to assess the environment and develop regulations according to its unique qualities, visual advantages, and important buildings. These should be taken as standpoints for design and planning. The design of features to go in a landscape or cityscape should be thoughtful and consider the particular characteristics of the setting, including color. Color is a powerful tool but is just one of many in a composition. Parameters regarding order, geometry, repetition, size, shape, and material should be part of the regulation criteria.

Reducing the variety and amount of elements in the visual field is crucial. Color may be protecting or decorating some element, which is obtrusive per se. The addition of features, such as signals and urban equipment, requires planning and restrictions in the color aspect too. In particular cases, the introduction of greenery may help to cover up the obtrusive elements, to create structure and order in the visual field or to balance a composition.

The adjustment of color dimensions (hue, saturation, and lightness) may be effective for resolving certain visual conflicts. Through the use of adequate colors, annoying features could be neutralized or kept inconspicuous. In this way, color would serve the purpose of contributing to visual order.

Cross-References

- [Achromatic Color](#)
- [Light Pollution](#)
- [Primary Colors](#)

References

1. Couto, M.: Contaminación visual del paisaje. https://sistemamid.com/panel/uploads/biblioteca/2014-04-30_02-21-0498509.pdf (2007). Accessed 2 Apr 2021
2. De Grandis, L.: *Theory and Use of Colour*. Harry N. Abrams, New York (1986)
3. Burga, J.: *Del espacio a la forma*. Facultad de Arquitectura, UNI, Lima (1987)
4. González, G. (ed.): *El gran libro del color*. Blume, Barcelona (1982)
5. Follis, J.: *Architectural Signing and Graphics*. Whitney Library of Design, New York (1979)
6. Wikipedia: Bird vision. http://en.wikipedia.org/wiki/Bird_vision. Accessed 10 Apr 2014

Color Pop-out

- [Color and Visual Search, Color Singletons](#)

Color Preference

Karen B. Schloss¹ and Stephen E. Palmer²

¹Department of Psychology and Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, WI, USA

²Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Synonyms

[Color aesthetics](#); [Color harmony](#); [Favorite colors](#)

Definition

How much people like different colors

Main Text

One of the most fascinating aspects of color cognition is that people have strong preferences, liking certain colors and color combinations much more than others. What is known about human

color preferences, not only in terms of *which* colors and color combinations people like but also *why* they like them, will be discussed below.

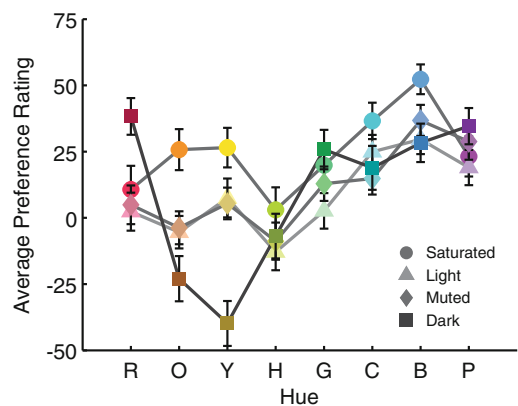
Preference for Single Colors

Average color preferences for a given sample of colors are typically measured behaviorally by asking a group of people to perform one of the following tasks [9]. First, observers can be asked to indicate which of two simultaneously presented colors they prefer for each possible pair of colors in the sample. The probability, averaged over observers, of choosing each color versus all other sample colors is then taken as a measure of its average relative preference within that sample. Second, observers can be shown all of the colors in the sample simultaneously and be asked to rank order them from most preferred to least preferred. The average rank of each color across observers provides a measure of average relative preference within the sample. Third, observers can be shown a single color on each trial and be asked to rate their preference for it on a discrete (e.g., 1–7) or continuous (e.g., marking along a line segment) rating scale. Average ratings across observers can then be taken as a measure of average relative preference for colors within the sample. Fourth, observers can be shown colors in a continuous space, or a sample of such colors, and be asked to select which color(s) they most like and most dislike. Responses are aggregated across observers to form frequency distributions of most liked and most disliked colors, which may be integrated (e.g., frequency liked minus frequency disliked) to arrive at overall preference measures or analyzed separately [5]. Color preference patterns using these different measures tend to be quite similar when the same observers (or large samples of different observers from the same population) judge the same colors.

Early researchers often claimed that color preferences were too idiosyncratic to be worth empirical study, but modern measurements of well-calibrated, computer-generated displays of standardized colors using improved data analysis techniques have now established that there are reliable and repeatable patterns in group data [17]. These average color preferences are most

easily understood in terms of the three primary dimensions of human color experience (see chapter ► “Color Spaces”): hue (its “basic color”), saturation (how vivid or pure the color is), and lightness or brightness (how light versus dark the color is). Figure 1 plots average adult preference ratings in the USA for 32 chromatic colors, consisting of eight hues – red, orange, yellow, chartreuse (yellow-green), green, cyan (blue-green) blue, and purple – in shades that are either highly saturated, light, muted (desaturated, mid-level lightness), or dark [8].

As Fig. 1 shows, average color preferences show a clear maximum around saturated blue and a clear minimum around dark yellow (greenish brown or olive). The majority of this variation is due to differences along the blue-to-yellow dimension of hue, with bluer colors being generally preferred to yellower colors. There is much less variation in the red-green dimension. In addition, people generally prefer more saturated colors over less saturated ones, with little difference between the light (pastel) and muted tones, at least in the USA. (See chapter ► “Color Preference”.) The most interesting finding theoretically is the striking difference between the shape of the hue preference curve for dark colors versus those for the saturation-matched light and muted colors. In particular, there are decreases in



Color Preference, Fig. 1 Color preference ratings in the USA as measured by Palmer and Schloss (2010). Preferences for the saturated, light, muted, and dark colors are plotted as a function of hue: red (R), orange (O), yellow (Y), chartreuse (H), green (G), cyan (C), blue (B), and purple (P)

preference for dark-orange (brown) and dark-yellow (greenish-brown or olive) relative to the light and muted oranges and yellows and modest increases in preference for dark-red and dark-green relative to the light and muted reds and greens [8]. The overall pattern of adult preferences for single colors is thus complex but clear and replicable. For a more extensive review of modern studies of single color preferences, the reader is referred to Whitfield and Wiltshire [17].

Describing Patterns of Color Preferences

Various types of models have been used to describe patterns of color preferences using different sets of predictors, typically computed using multiple linear regression. These models provide a parsimonious way to describe patterns of color preferences of groups or individuals, simply by specifying the weights placed on each of the predictors. One series of models used predictors that correspond to different stages of visual processing. The first was the *cone contrast model*, which used a weighted average of cone contrasts relative to the background. Cone contrasts are believed to be computed very early in visual processing: $L-M$ and $S-(L + M)$, where S , M , and L represent the outputs of cones maximally sensitive to short, medium, and long wavelengths of light, and was later extended to include luminance ($S + M + L$) and saturation [4]. The *color appearance model* tested higher-level color appearance dimensions, using a weighted average of observer-rated redness-greenness, blueness-yellowness, saturation, and lightness of each color, roughly analogous to their coordinates in the natural color space (see chapter ► “[Psychological Color Space and Color Terms](#)”) [8]. A third type of model tested perceptual dimensions cylindrical coordinates in CIELAB color space, using a linear combination of the 1st harmonic over hue angle (reflecting preference for a dominant hue, such as blue), a 2nd harmonic over hue angle (reflecting preference for both ends of a dominant opponent axis, such as red and green), lightness, and chroma [10]. Comparisons across model types demonstrate that higher-level predictors (e.g., color appearance and cylindrical coordinates) better describe color preferences than

lower-level predictors (e.g., cone contrast); see [10] for a review and more details on modeling color preferences using color space metrics.

Another type of model aimed at describing color preferences used *color emotions*, which are associations between emotion-related terms and colors [7]. A large set of color emotions can be reduced using factor analysis, producing factors such as *active-passive*, *light-heavy*, and *cool-warm*. Those factors are used to describe color preferences using linear regression techniques [11]. Evidence suggests that models using color emotions describe color preferences nearly as well as color appearance dimensions do [2].

Explaining Color Preferences

This far, this entry has focused on describing *which* colors people like and dislike. But *why* do they like the ones they like? Indeed, why do people have color preferences at all? One theory that is that color preferences are hardwired into the visual system to serve an adaptive function on an evolutionary time scale [3]. However, evidence demonstrating individual differences and temporal changes in color preferences suggests that a large aspect of color preferences is learned through experiences (see [12] for a review). These differences, as well as the general pattern of color preferences in Fig. 1, can be explained by the ecological valence theory (EVT). Originally, the EVT posited that people like colors to the degree that they like the environmental objects that are characteristically those colors [8] but has been extended to include all concepts associated with colors (e.g., objects, entities, and events) [12]. For example, people tend to like blues and cyans because they like blue things on average (e.g., clear sky and clean water) and they tend to dislike browns and olive colors because they dislike brown and olive things on average (e.g., biological waste and rotting food). The theoretical rationale of the EVT is that it will be adaptive for organisms to approach objects and situations associated with the colors they like and to avoid objects and situations associated with the colors they dislike to the extent that their color preferences are correlated with objects and situations that they like versus those that they do not like.

This does not mean that color preferences are irrelevant to object preferences: clearly they matter for functionally identical artifacts that are available in many colors (e.g., clothes, furniture, cars, and appliances). However, the EVT suggests that those color preferences arose initially from associations with characteristically colored objects and were then positively or negatively reinforced by positive or negative experiences associated with those colors.

Support for the EVT was first provided through empirical measurements of the weighted affective valence estimates (WAVEs) for the 32 chromatic colors in Fig. 1. The WAVE for each color measures the extent to which people like the objects that are associated with that color. It was computed from observers' average valence (liking/disliking) ratings for *all* things named as associates for each color studied, with each valence rating being weighted by the similarity of the given color to the color of that associate. Although blue is strongly associated with objects that are almost all liked (e.g., clear sky, clean water, swimming pools, and sapphires), most colors have associates with both positive and negative valences. Nevertheless, the WAVE, computed as specified above, explained 80% of the variance in the data shown in Fig. 1 with no estimated parameters. Further support came from studies showing that individual differences in color preferences could be accounted for by individual differences in preferences for objects associated with colors. Moreover, studies using a controlled priming task and natural variations over seasons demonstrated that color preferences systematically changed over time, depending which concepts were activated in people's minds. For example, priming people to think about ripe strawberries and roses (positive in valence) led to an increased preference for red. Both individual and temporal differences can be understood within a single theoretical framework [12].

Development of Color Preferences

Infant color preferences are studied by examining the looking behavior of babies when they are shown pairs of colors side by side. It is generally

assumed that the color at which the baby looks longer and/or first is preferred to the alternative. Color preferences are therefore measured by determining the average looking times and/or the probabilities of first-looks [15]. Infants younger than about 3–4 months tend not to be studied because the short-wavelength sensitive cones do not mature until that age, making them functionally color deficient relative to adults (see chapter ▶ “Effects of Color Terms on Color Perception and Cognition”).

When infant looking preferences are measured for highly saturated colors, the hue preference function tends to have roughly the same shape as the corresponding adult hue preference function, with a peak around blue and a trough around yellow-to-yellow-green [5]. Great care has to be taken to ensure that hue-based color preferences actually reflect differences in hue by controlling for luminance, brightness, discriminability, colorimetric purity, and saturation. More recent studies that directly compared infant preferences with those of adults for less saturated, but better matched, colors have found important differences, however. In particular, infant preferences for these color samples vary primarily on the red-to-green dimension, with redder colors being more preferred, and do not vary much on the blue-to-yellow dimension [2]. Because this pattern for infants is opposite that for adults, color preferences either must be subject to learning as a result of experiences with differently colored objects or there must be a substantial maturational process that influences color preferences.

As children develop, gender differences begin to emerge. For example, 1-year-old children show no difference in preference for pinks vs. blues, but around 2, girls tend to prefer pinks more than boys do [6]. This has been explained by the corresponding gender differentiation in toys, where toys marketed for girls tend to come in pinks and toys marketed for boys tend to come in blues. Thus, this gender differentiation in color preferences can be explained by the EVT, within which children come to like colors associated with objects they like. The role of gender in adult color preferences is still an open question [4].

Color Preferences in Different Contexts

A question of applied interest is how adult color preferences for patches of “context free” colors, as described above, generalize to preferences for colored objects. Presumably, they would not generalize for natural objects that have prototypical colors (e.g., not preferring bananas in blue, given their natural colors are yellows/greens), but generalization is evident for artifacts that could, in principle, be produced in virtually any color, such as shirts, walls, sofas, and cars [14]. Some have shown that hue preferences are fairly consistent across different types of objects, with differences occurring primarily in lightness and saturation [14]. For example, people generally preferred lighter walls and darker couches and disliked saturated colors for most contexts even though saturated colors were most preferred for contextless colored patches. These variations seem to depend on practical considerations, such as walls being preferred in lighter tones that make rooms appear more spacious and couches being preferred in darker shades that hide dirt. However, others have suggested that hue preferences are more variable across contexts [5], and more research is needed to fully understand how and why color preferences vary with context.

Preference for Color Combinations

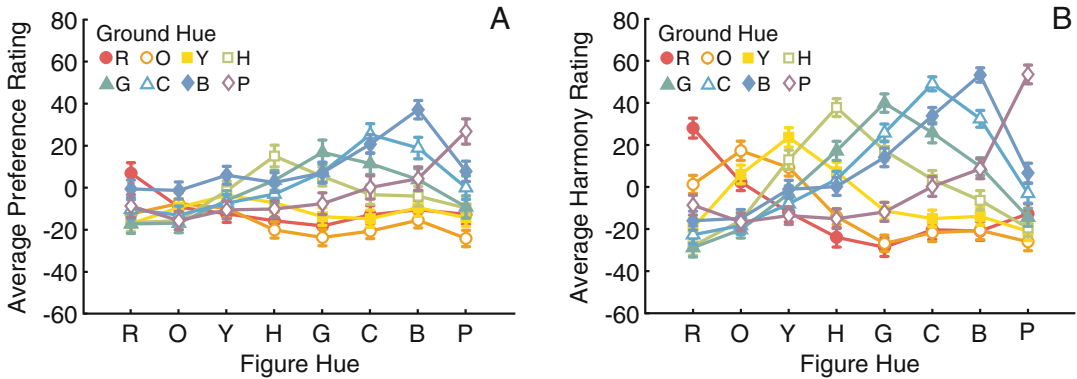
Chevreul formulated a highly influential art-based theory of color harmony (or color preference, because he used the terms interchangeably), which claimed that there are two distinct types: *harmony of analogous colors* and *harmony of contrast* [1]. In brief, harmony of analogous colors includes *harmony of scale* (colors of the same hue that are similar in lightness) and *harmony of hues* (colors that are similar in hue and the same in lightness). Harmony of contrast includes *harmony of contrast of scale* (colors of the same hue that differ in lightness), *harmony of contrast of hues* (colors of similar hue that differ in lightness), and *harmony of contrast of colors* (colors that differ in both hue and lightness). Other theories include Itten’s claim that combinations of colors are harmonious provided that the colors produce neutral gray when mixed as paints, Munsell’s and Ostwald’s theories that colors are

harmonious when they have certain relations in color space (e.g., they are constant in hue and saturation but vary in lightness), as well as other theories proposed by Goethe and Moon and Spencer [16]. None of these theories was formulated on the basis of aesthetic measurements, although some have since been tested empirically.

Empirical Research on Color Pair Preference and Harmony in Combinations

The confusion surrounding preferences for color pairs can be clarified by distinguishing among three different concepts: *pair preference*, *pair harmony*, and *figural preference*, for a color against a colored background [11]. *Pair preference* is how much an observer *likes* the combination of the two colors as a whole. *Pair harmony* refers to how well the two colors *go together*, regardless of whether the observer likes the combination or not. This distinction between preference and harmony can be understood by analogy to music, wherein nearly everyone would agree that Mozart’s music is more harmonious than Stravinsky’s, but some prefer Mozart, and others prefer Stravinsky. *Figural preference* denotes how much the observer likes the single color of the figure when viewed against a different color in the background. Although figural preference involves a judgment about the single color of the figure, it is relevant to preferences for color combinations because the same color can look strikingly different on different background colors (see chapter ► “Simultaneous Color Contrast”).

Figure 2A shows average preference ratings for color pairs as a function of the hue of the figure (x-axis) and that of the ground (the different curves). The primary pattern in the data is that, for every background hue, people prefer combinations in which the figure has the same or a very similar hue. On average, people tend to like color combinations that are the same or similar in hue but differ in lightness, which Chevreul called harmonies of analogous colors. There is no evidence for Chevreul’s harmonies of contrast, however, because there are not reliable increases in the functions at opposite hues (e.g., red and green). A secondary fact is that people tend to like color combinations to a degree that reflects their



Color Preference, Fig. 2 Rated preference (a) and harmony (b) for pairs of colors as a function of the figural hue (x-axis) and ground hue (different curves)

preferences for figure and ground colors, with combinations on blue backgrounds being generally most preferred and those on orange backgrounds being least preferred. This tendency is relatively minor, however, accounting for only 22% of the variance in preference ratings.

Harmony ratings of the same color pairs are plotted in Fig. 2B. They are similar to preference ratings, but the peaks at same hue combinations are even more pronounced. This similarity explains why preference and harmony have so often been equated, but there are differences between the two measures. Pair preference depends more on preference for the component colors within the pair, and harmony depends more on color similarity.

Ratings of figural preference for colors against colored backgrounds are measurably different from ratings of both pair preference and pair harmony, showing clear effects due to the hue contrast and lightness contrast of the figure against the background: warm figures (e.g., red, orange, and yellow) were preferred against cool backgrounds (e.g., green, cyan, and blue) and vice versa [14]. It appears that what Chevreul termed harmonies of contrast actually apply to judgments about preferences for figural colors when viewed against different colored backgrounds. Given the general preference for saturated colors described previously, it is not surprising that observers prefer

figural colors against highly contrastive background hues, because these would produce the strongest simultaneous color contrast effects, thus increasing the perceived saturation of the figural region.

Theories of Preferences for Color Combinations

The foregoing describes which color combinations people prefer and which ones they find harmonious, but why do these variations in preference and harmony arise? To the extent that pair preferences are influenced by preferences for the component colors, ecological associations of colors with objects are one important factor. Pair preferences are also influenced by people's positive/negative associations with objects and/or institutions that are associated with those colors in combination (e.g., red and green with Christmas). In a related vein, Schloss, Poggesi, and Palmer [13] investigated preferences for school colors among Berkeley and Stanford students: blue and gold for Berkeley and red and white for Stanford. They found that Berkeley students liked Berkeley color combinations better than Stanford students did, and Stanford students like Stanford colors better than Berkeley students did, with these effects increasing with increasing amounts of self-rated school spirit. Such results clearly imply that ecological effects are present in preference

for color combinations as well as for individual colors.

Perhaps the most important finding about preferences for color pairs is that people generally like harmonious combinations of the same (or similar) hue that differ in lightness. Although it is not immediately obvious why this might be the case from an ecological viewpoint, color harmony might stem from ecological color statistics in natural images corresponding to different areas of the same ecological object [14]. A red sweatshirt, for example, would be darker red where it was in shadow and lighter red where it was brightly illuminated. Accordingly, pairs that are judged to be most harmonious (i.e., that “go together” best) may, in fact, be those that are most likely to co-occur within the same object in natural images.

Cross-References

- [Effects of Color Terms on Color Perception and Cognition](#)
- [Psychological Color Space and Color Terms](#)
- [Simultaneous Color Contrast](#)

References

1. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors*. Van Nostrand Reinhold, New York (1839/1967)
2. Franklin, A., Bevis, L., Ling, Y., Hurlbert, A.: Biological components of colour preference in infancy. *Dev. Sci.* **13**, 346–354 (2009)
3. Hurlbert, A.C., Ling, Y.: Biological components of sex differences in color preference. *Curr. Biol.* **17**, 623–625 (2007)
4. Hurlbert, A.C., Owen, K.A.: Biological, cultural, and developmental influences on color preference. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) *Handbook of Color Psychology*, pp. 454–480. Cambridge University Press, Cambridge (2015)
5. Jonauskaitė, D., Mohr, C., Antonietti, J.-P., Spiers, P.M., Althaus, B., Anil, S., Dael, N.: Most and least preferred colours differ according to object context: new insights from an unrestricted colour range. *PLoS One*. **11**, e0152194 (2016)
6. LoBue, V., DeLoache, J.S.: Pretty in pink: the early development of gender-stereotyped colour preferences. *Brit. J. Dev. Psychol.* **29**(3), 656–667 (2011)
7. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modeling. *Color. Res. Appl.* **29**, 381–389 (2004)
8. Palmer, S.E., Schloss, K.B.: An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8877–8882 (2010)
9. Palmer, S.E., Schloss, K.B., Sammartino, J.: Visual aesthetics and human preference. *Annu. Rev. Psychol.* **64**, 77–107 (2013)
10. Schloss, K.B., Lessard, L., Racey, C., Hurlbert, A.C.: Modeling color preference using color space metrics. *Vis. Res.* **151**, 99–116 (2018)
11. Schloss, K.B., Palmer, S.E.: Aesthetic response to color combinations: preference, harmony, and similarity. *Atten. Percept. Psycho.* **73**, 551–571 (2011)
12. Schloss, K.B., Palmer, S.E.: An ecological framework for temporal and individual differences in color preferences. *Vis. Res.* **141**, 95–108 (2017)
13. Schloss, K.B., Poggesi, R.M., Palmer, S.E.: Effects of university affiliation and “school spirit” on color preferences: Berkeley versus Stanford. *Psychon. B. Rev.* **18**, 498–504 (2011)
14. Schloss, K.B., Strauss, E.D., Palmer, S.E.: Object color preferences. *Color. Res. Appl.* **38**, 393–411 (2013)
15. Teller, D.Y., Civan, A., Bronson-Castain, K.: Infants' spontaneous color preferences are not due to adult-like brightness variations. *Vis. Neurosci.* **21**, 397–402 (2004)
16. Westland, S., Laycock, K., Cheung, V., Henry, P., Mahyar, F.: Colour harmony. *Colour: Des Creat.* **1**, 1–15 (2007)
17. Whitfield, T., Wiltshire, T.: Color psychology: a critical review. *Genet. Soc. Gen. Psychol. Monogr.* **116**, 387–411 (1990)

Color Processing in the Inferior Temporal Cortex

- [Color Information Processing in Higher Brain Areas](#)

Color Processing in the Ventral Higher Visual Area

- [Color Information Processing in Higher Brain Areas](#)

Color Processing, Cortical

Daniel C. Kiper

Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Synonyms

Chromatic processing; Cortical

Definition

The transformation of color signals and chromatic properties of receptive fields within the visual cortex of primates.

Processing of Chromatic Signals in the Early Visual Pathways

The processing of chromatic signals in the retina and lateral geniculate nucleus (LGN) has been the focus of numerous studies and is well understood. Less is known about the fate of color signals in the cortex. This entry first reviews central aspects of color processing in the primary visual cortex (V1) and discusses how it differs from subcortical processes. It then discusses the processing of color signals in extrastriate visual areas.

Color in the Striate Cortex (V1)

Chromatic Properties of Individual Neurons

In the primate primary visual cortex, it had been estimated that about 50% of the cell population is selective for color [1]. Estimates of the proportion of color-selective cells, however, are complicated by the use of different criteria for the classification across studies. Interestingly, color signals in primary visual cortex have long been thought to be relatively weak relative to black and white stimuli. Studies based on imaging techniques of the active human brain, however, have now clearly

demonstrated that primary visual cortex contains a large proportion of color-responsive and color-selective cells [2]. The majority of color-selective cells in V1, like those of its subcortical input, the lateral geniculate nucleus (LGN), simply add or subtract their chromatic inputs. Indeed, most V1 [3] cells' responses to chromatic modulations are well accounted for by a model postulating a linear combination of the signals derived from the three cone classes. Although there are some V1 cells that are more selective for color than predicted [4], the model adequately fits the responses of the majority of V1 cells.

The preferred colors of V1 neurons [2], however, do not cluster around the three cardinal directions found in the LGN [5]. Instead, cells often prefer colors that lie intermediate to these "cardinal" directions. For cortical cells, the classification of chromatic cells into red-green or blue-yellow opponent cells is therefore not valid anymore.

Moreover, most color-selective cells in V1 [6] respond also vigorously to luminance variations, a property that seems ubiquitous in the visual cortex. It also illustrates the fact that color selectivity does not imply color opponency. A response to stimuli containing chromatic but no luminance information (isoluminant stimuli) does not imply that the neuron receives opponent inputs from two or three cone classes. True color opponency can be deduced in neurons that give stronger responses to isoluminant than to luminance stimuli, provided that the stimuli are equated for cone contrasts [7]. It thus appears that in the cortex, there is a whole continuum of cells, ranging from strict color opponency to strict luminance [8].

Recent studies have also shown that, unlike earlier levels, primary visual cortex contains so-called double-opponent cells [6]. Double-opponent cells are cells whose receptive field combines color and spatial opponency. The defining functional property of these cells is that they respond strongly to color patterns but weakly or not at all to uniform (full-field) color modulations. The existence of pure double-opponent cells, cells with non-oriented receptive fields that respond only to color patterns [9], has been hotly debated in the last years. While early reports have been

contested, it now appears that double opponency does exist in primary visual cortex but linked with other functional properties such as orientation selectivity and a high sensitivity for luminance stimuli as well [6].

Clustering and Specialization of Color-Selective Cells in V1

In the last decades, the localization and functional specialization of color-selective cells within area V1 have been the matter of debate. A number of studies reported that color-selective cells in V1 are not orientation selective (a canonical property of virtually all primate V1 neurons) and that these cells are clustered within the cytochrome oxidase (CO)-rich blobs that have been described previously [10]. Both claims, however, have been contested by a number of anatomical and physiological studies [8]. Some of the controversy can be attributed to the different techniques and stimuli that have been used to localize and characterize color-selective cells. A number of extensive and careful studies using single-cell recordings combined with histological processing have clearly shown that color-selective cells are not restricted to the CO blobs.

The same studies have shown that color selectivity and orientation selectivity are not, as originally proposed, segregated within the V1 cell population. The most convincing, but not unique, evidence for conjoint selectivity to color and orientation comes also from single-cell recordings [11].

The Case of Blue-Yellow Signals

For a number of mostly technical reasons, many color-vision studies have focused primarily on red-green opponency and less on the evolutionary older blue-yellow system. The discovery of the koniocellular layers in the LGN and their predominant role in the processing of blue cone signals (necessary for blue-yellow opponency) [12] raised the possibility of a specialized treatment of blue cone signals within the primary visual cortex. To what extent blue cone signals are clustered within V1 and how much they contribute to cortical color processing is not clear yet. Recent studies seem to indicate that blue cone signals are

distributed uniformly within V1 (no clustering) and that these signals might be combined with achromatic signals in the double-opponent cells described above.

Color in the Extrastriate Cortex

Although several studies showed that individual neurons in the dorsal visual pathway, in particular in area MT of the macaque monkey, can significantly respond to chromatic variations [13], these responses are typically smaller than those obtained with luminance stimuli and do not account for the animal's behavioral performance. The following sections are thus restricted to areas of the ventral pathway that are known to play a critical role in color processing.

Proportion of Color-Selective Cells

In extrastriate areas of the ventral pathway, the number of neurons whose responses are affected by color variations remains surprisingly constant despite the variability of the criteria used to classify neurons. This proportion reaches 50% in area V2 and 54% in V3 [1]. Estimates in later areas of the ventral pathways (reviewed in 8) are more variable. In area V4, often but mistakenly considered a "color" area of the primate brain, original estimates ranged from less than 20–100%. A more recent estimate of 66% has been reported. In the IT cortex, it has been estimated around 48%.

Chromatic Properties of Individual Neurons

As already described in V1 neurons, the chromatic properties of extrastriate neurons differ from those in the retina and LGN in two important ways. First, a significant proportion of neurons in each area possess a high degree of color selectivity. These neurons show a narrow tuning in color space. Narrowly tuned neurons have been reported, in different proportions, within areas V2, V3, and V4. The second major difference already described in V1 concerns individual neurons' preferred colors. As in V1, the distribution of preferred colors of extrastriate neurons is not clustered in color space but is uniformly distributed [8].

Color Versus Other Visual Attributes

The question of conjoint selectivity for color and other visual attributes has been posed and answered for V1 neurons (see above). Similarly, the question whether color and other visual attributes such as orientation, motion, and size are segregated in the extrastriate cortex has been studied. Numerous studies (see [8]) in the last decade have shown that cells within V2, V3, and V4 can be concomitantly tuned for several dimensions of the visual stimulus. It thus appears that color is not processed independently but by the same neuronal populations that also code orientation or size. Note however that a recent study reported the existence of a significant subpopulation of V4 cells that respond to chromatic but not luminance variations leading the authors to suggest that color and luminance might be treated by different channels within area V4 [14].

Clustering of Color-Selective Cells

The debate concerning the organization of color-selective cells into clusters within a given area has extended to several visual areas beyond V1. In V2, the thin bands defined by CO staining have been reported to represent clusters of color-selective cells and to be the source of the color signals sent to area V4 and areas of the inferotemporal cortex. Moreover, optical imaging studies of area V2 concluded that color is represented in an orderly fashion within the thin stripes in the form of well-defined color maps that resemble those based on human color perception. As in V1, however, the clustering of color selectivity within the thin CO stripes has been challenged.

In areas more posterior within the temporal pathway, color-selective cells have been reported to cluster into subregions. This clustering is in fact often offered as an explanation for the large variance in estimates of the proportion of color-selective cells in the temporal cortex. Microelectrode recordings in dorsal V4 are thought to encounter many color clusters that seem to be much less prevalent in ventral parts of V4.

In the most anterior parts of the ventral pathways, the distinction between cortical areas is

much less clear, and the relationship between areas reported in the macaque brain with those of the human is still controversial. It has nonetheless been suggested that color might be treated within a specialized pathway that extends across several of the ventral visual areas including V2, V4, and the dorsal portion of the posterior inferotemporal cortex (PITC) [15]. Within PITC, color-selective cells would be clustered into islands themselves containing orderly, columnar color maps, reminiscent of the organization previously reported in V2.

Is There a Color Center in the Primate Brain?

Several of the issues discussed above are intimately related to the question whether the primate brain contains one color center, a cortical area whose main function would be to support most or all aspects of color perception. The notion of a color center is a natural consequence for the proponents of a strictly modular view of cortical organization. Most researchers agree that there are areas in the ventral cortex that are highly activated by color stimuli, but none of these areas have been established as predominant over the others.

Moreover, experimental lesions in macaque V4, the most popular candidate for the role of color center, do not result in a complete loss of color vision. These results cast doubts on the notion of a unique color center and support the idea that color, like many other visual attributes, is treated within a network of neuronal populations distributed within the ventral occipitotemporal pathway.

Unresolved Issues

Important perceptual phenomena such as color constancy, unique hues, or color categorization remain largely unexplained. Further studies reconciling imaging or lesion results in the human brain with those obtained by methods revealing single activity in the macaque brain are thus necessary to fill these gaps. Moreover, the relationship between color signals and those associated with other visual attributes such as object shape or

motion needs further scrutiny, particularly in extrastriate areas. Today, it is safe to say that a full understanding at the neuronal level of perceptual phenomena associated with color is still eluding the vision science community.

Cross-References

► Unique Hues

References

1. Zeki, S.: The distribution of wavelength and orientation selective cells in different areas of the monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* **217**, 449–470 (1983)
2. Engel, S., Zhang, X., Wandell, B.: Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*. **388**(6637), 68–71 (1997)
3. Lennie, P., Krauskopf, J., Sclar, G.: Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**(2), 649–669 (1990)
4. Cottaris, N.P., De Valois, R.L.: Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature*. **395**(6705), 896–900 (1998)
5. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic properties of neurons in macaque LGN. *J. Physiol.* **357**, 241–265 (1984)
6. Johnson, E.N., Hawken, M.J., Shapley, R.: The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci.* **4**, 409–416 (2001)
7. Schluppeck, D., Engel, S.A.: Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording. *J. Vis.* **2**(6), 480–492 (2002)
8. Gegenfurtner, K.R., Kiper, D.C.: Color vision. *Annu. Rev. Neurosci.* **26**, 181–206 (2003)
9. Shapley, R.M., Hawken, M.J.: Color in the cortex: single- and double-opponent cells. *Vis. Res.* **51**, 701–717 (2011)
10. Conway, B.R., Chatterjee, S., Field, G.D., Horwitz, G.D., Johnson, E.N., Koida, K., Mancuso, K.: Advances in color science: from retina to behavior. *J. Neurosci.* **30**(45), 14955–14963 (2010)
11. Leventhal, A.G., Thompson, K.G., Liu, D., Zhou, Y., Ault, S.J.: Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* **15**(3), 1808–1818 (1995)
12. Henry, S.H., Reid, R.C.: The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* **23**, 127–153 (2000)
13. Gegenfurtner, K.R., Kiper, D.C., Beusmans, J.M., Carandini, M., Zaidi, Q., Movshon, J.A.: Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* **11**(3), 455–466 (1994)
14. Bushnell, B.N., Harding, P.J., Kosai, Y., Bair, W., Pasupathy, A.: Equiluminance cells in visual cortical area v4. *J. Neurosci.* **31**(35), 12398–12412 (2011)
15. Conway, B.R., Tsao, D.Y.: Color architecture in alert macaque cortex revealed by fMRI. *Cereb. Cortex.* **16**(11), 1604–1613 (2006)

Color Psychology

Zena O'Connor

Design Research Associates, University of Sydney, Sydney, NSW, Australia

Synonyms

[Chromotherapy](#); [Psychological impact of color](#)

Definition

Color psychology refers to a branch of study which postulates that color has a range of psychological or behavioral responses.

Overview

Color psychology and color therapy exist on the periphery of alternative medicine and are generally not accepted under the auspices of mainstream medical science and psychology. Despite this, a plethora of articles can be found in mainstream and digital media that discuss a link between color and a range of psychological, cognitive, biological, and behavioral effects. While it is often promoted that such a link exists on a universal, causal basis by some sectors of the media, there is minimal evidence to support this hypothesis. Furthermore, there are a number of factors that influence color and human response, and these include individual differences (such as age, gender, affective state, belief systems, and

environmental stimuli screening ability), social and cultural differences, as well as contextual and temporal differences. Despite this, claims such as *red is stimulating and arousing* and *blue is calming, relaxing, and healing* are often quoted without substantiation or evidence of any nature in popular media. The source of such claims can be traced to a number of key authors including Faber Birren, Kurt Goldstein, Robert Gerard, and Max Lüscher, whose works have been either superseded or debunked for lack of scientific rigor. However, some proponents have a vested interest in promoting a link between color and human response due to the popularity of such claims and the opportunity to capitalize on this popularity through the sale of books, workshops, courses, and consultancy services. While there is a place for such products and services, there are also compelling reasons to apply the principle of caveat emptor to color psychology and color therapy claims found in nonacademic sources.

Defining Color Psychology

The term “color psychology” suggests that the interface between color and human response is underpinned by a causal link wherein color has the capacity to influence a range of human responses including affective, cognitive, and related behavioral responses. In mainstream media and popular culture, “the psychological effects of color” is assumed to be a causal link that exists on a universal basis irrespective of individual differences or the impact of cultural, contextual, or temporal factors [1, p. 9]. This understanding about color psychology remains fairly consistent; however, some sources extend the definition to include a larger gamut of responses from ► [color preference](#) to precognitive, psychophysical, and “biology-based responses,” defined as responses on human metabolism, circulation, and respiratory systems [2, p. 92].

The word “therapy” as in “color therapy” generally refers to a remediation of a health or psychological issue intended to address a diagnosed problem or stop a medical or psychological condition from occurring. Hence, color therapy implies that color can be used as a therapeutic

device in the remediation of a health or psychological issue. Some writers take this concept further and suggest that the “biological consequences of color responses can be a valuable tool in health management” for the treatment of various ailments, and this appears to be a fairly common understanding of the term [2, p. 93], while other writers suggest that color therapy, or chromotherapy, can be used prescriptively as a “truly holistic, non-invasive and powerful therapy which dates back thousands of years” [3, p. 1].

The Interface Between Color and Human Response: Recent Research

Academic literature includes a broad range of studies that discuss the effects of light as well as the influence of colored light waves in respect to human response. While human vision is a complex and not yet fully understood process, the receptor system for detecting light has been found to be different from that associated with the circadian cycle [4]. In reference to human circadian cycles, light tends to regulate our sleeping patterns and changes in light–dark exposure can desynchronize the circadian cycle affecting the ability to sleep and wake, as well as impacting on physiological and metabolic processes. Disruptions to the circadian rhythm may result in changes in mood and behavior as evidenced by studies that focus on seasonal affective disorder, known by the acronym SAD (seasonal affective disorder) [5–7]. Light has also been found to have an effect on the human neuroendocrine system and may also suppress melatonin and elevate cortisol production, which can have negative impacts [7, 8]. Furthermore, a number of recent studies have indicated that certain wavelengths of light may have specific impacts. For example, blue light may improve cognitive performance; different colored lenses may assist with reading difficulties such as dyslexia; and the human circadian system may be particularly sensitive to short-wavelength light [9–11]. Despite the many advances in recent research, the precise roles of the rods and cones of the retina as well as melanopsin in the control of circadian cycles remain to be determined [12].

A plethora of studies exist which suggest that color may influence a range of psychological, physiological, and behavioral responses; however, the range and diversity of research findings was highlighted by an analysis of 30 studies [13]. For example, it has been suggested that red has a greater capacity for arousal than blue [14–16]. However, findings from a more recent study suggest that there is no statistically significant difference between these two colors in terms of physiological arousal and that it may be hue rather than saturation (intensity) of color that has an impact [17]. In addition, recent studies have found that responses to color may vary depending on age, gender, culture, and preference [18, 19]. It is important to note that while many of the recent studies that focus on psychological, physiological, and behavioral responses are scientifically rigorous, the findings are often based on an extremely limited range of color samples or a small sample group. In terms of color therapy, while associations may exist between various colors and a range of different human responses, it does not necessarily follow that such colors can be effectively used in therapy as some sources in popular media suggest.

Color Psychology in Popular Media

Information about color psychology in popular media is abundant and accessible via myriad magazines, online magazines, and Internet websites. Internet search engine Google currently provides access to 365,000 websites for information relating to color psychology. For example, www.colorthrapyhealing.com provides numerous pages on color psychology and color healing, respectively, including fairly detailed information about the use of color as a therapeutic tool and a range of color therapy workshops. Similarly, www.colour-affects.co.uk offers detailed comment about the so-called psychological properties of color and provides a summary of four personality types supposedly linked to specific colors. In addition, mass media magazines and online magazines such as *WellBeing* (www.wellbeing.com.au) feature various articles on color psychology and color therapy.

The information available from popular, mainstream media can vary from broad, generalized commentary to detailed, pseudoscientific discussions peppered with relatively harmless generalizations, platitudes, and motherhood statements, at best, and unsubstantiated and highly dubious claims, at worst, such as “colors are the mother tongue of the subconscious” and “color heals.”

Some sources of information provide claims about perceptual color effects and color associations intermingled with unsubstantiated claims about color psychology such as:

Although red, yellow and orange are in general considered high-arousal colours and blue, green and most violets are low-arousal hues, the brilliance, darkness and lightness of a colour can alter the psychological message. While a light blue-green appears to be tranquil, wet and cool, a brilliant turquoise, often associated with a lush tropical ocean setting, will be more exciting to the eye. The psychological association of a colour is often more meaningful than the visual experience.

Colours act upon the body as well as the mind. Red has been shown to stimulate the senses and raise the blood pressure, while blue has the opposite effect and calms the mind.

People will actually gamble more and make riskier bets when seated under a red light as opposed to a blue light. That's why Las Vegas is the city of red neon. [20]

Other sources of information suggest that color has therapeutic effects such as the claims provided in the online magazine *Conscious Living*.

Colour can repair and heal the body, when the frequency of the colour aligns with the emotion needed to activate the micro-particulars so healing can take place. . . The use of colour in visualisation is most effective, and easiest for the novice to utilise, as colour has a very strong radiating effect on the whole body. Every other form of colour therapy is fundamentally symbolic. [21]

In a similar vein, Wright suggests that

In practice, colour psychology works on two levels: the first level is the fundamental psychological properties of the eleven basic colours, which are universal, regardless of which particular shade, tone or tint of it you are using. Each of them has potentially positive or negative psychological effects and which of these effects is created depends on personality types and —crucially— the relationships within colour combinations, the second level of colour psychology. [22]

Wright offers further clarification from the colour affects system, and it should be noted that the colour affects site offers color consultancy services plus a range of industry and personal courses and workshops.

Similarly, Rewell provides a detailed discussion about specific responses to color in the online magazine *WellBeing* as follows:

Babies cry more in yellow rooms. Tension increases in people in yellow rooms and people who drive yellow cars are more prone to become aggravating in heavy traffic... Spend time exposed to a lot of yellow and you'll feel like time has sped up... A rejection of yellow indicates a fear of change... Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration. [23]

Other sources of information about color psychology include architectural and interior design books (e.g., see Kopacz and Mahnke) as well as technical reports [2, 24, 25]. In addition, various courses and workshops discuss issues relating to color psychology including the Colour Therapy Healing organization in the UK, the International School of Colour and Design in Sydney, and the Nature Care College in Sydney [26–28].

Information from sources such as those discussed above often quotes a range of color associations co-mingled with various psychological, cognitive, or behavioral responses to color. In addition, such sources rarely if ever provide evidence to substantiate any claims regarding the impact of color on human response. An example of the intermingling of a range of different types of response is provided by Cherry as follows:

Red is a bright, warm color that evokes strong emotions. Red is also considered an intense, or even angry, color that creates feelings of excitement or intensity.

Blue calls to mind feelings of calmness or serenity. It is often described as peaceful, tranquil, secure, and orderly. Blue can also create feelings of sadness or aloofness. Blue is often used to decorate offices because research has shown that people are more productive in blue rooms.

Green is restful, soothing, cheerful and health-giving. Green is thought to relieve stress and help heal. Those who have a green work environment experience fewer stomach aches. Green has long been a symbol of fertility. [1]

Some sources go so far as to state symbolic color associations and then extend these further to include different situations or contexts such as the following:

Red is the colour for courage, strength and pioneering spirit... It is the colour of anger, violence and brutality.

Blue is calming, relaxing and healing (but) not as sedentary as indigo.

Green is the colour of balance and harmony and can, therefore, be helpful in times of stress. If one has experienced trauma, a green silk wrapped around the shoulders can have a very therapeutic effect. [3]

It is not uncommon for claims such as those above to be presented in an authoritative manner, exhorting the reader to believe and accept; however, the lack of evidence indicates that such claims are often personal opinion masquerading as scientific fact or pseudoscience.

An Irrefutable, Causal, and Universal Link Between Color and Human Response?

Various sources of information in mainstream media and related sources tend to suggest that an irrefutable, causal relationship exists between color and human response, as illustrated by Hill's statement: "Based on numerous studies by Drs Morton Walker, Gerard and Faber Birren, the link between color and physiological responses has been well documented" [24, p. 7].

In addition, many sources imply that the link between color and human response is universal irrespective of individual or cultural differences. For example, Logan-Clarke asserts: "Red... is stimulating and energizing therefore it is helpful for tiredness and lethargy, to stimulate low blood pressure, to boost sluggish circulation... Red is energizing and excites the emotions, and can stimulate the appetite" [3, p. 10]. Similarly, Rewell contends "Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration" [23], while Kopacz suggests "Red is believed to sensitise the taste buds and sense of smell, increasing the appetite... all this occurs because the heart rate instinctively quickens, which causes a release of adrenalin into the bloodstream raising blood pressure and stimulating the nerves" and "the sight of the color blue causes the

body to release tranquilizing hormones when it is surveyed, particularly a strong blue sky” and “many believe (blue) can lower blood pressure, slow the pulse rate and decrease body temperature” [2, pp. 76, 79].

In regard to such claims, Kopacz notes the lack of evidential support but offers the work of a number of authors and designers including Birren and Mahnke as well as Morton Walker (author of *Bald No More, Foods for Fabulous Sex, Your Guide to Foot Health*, as well as *The Power of Color*) and Wright (color consultant of www.colour-affects.co.uk) to support his color psychology claims. Similarly, Mahnke notes the inconclusive nature of findings from research that focuses on the interface between color and human response, but nevertheless refers to the work of Birren and Goldstein among others as support for various psychophysiological effects of color.

In regard to color therapy, it is suggested that color can be used as a treatment tool in conjunction with the seven chakras of the body [2, 3, 25]. The concept of chakras, considered to be energy centers within the human body, belongs to a belief system originating in the Hindu scriptures known as the Upanishads, dating from the first millennium BCE. Under the color chakra theory, a color is linked to each of the seven chakra and these colors are associated with bodily functions and dysfunctions within each chakra area. For example:

Red: Activates the circulation system and benefits the five senses; used to treat colds, paralysis, anaemia, ailments of the bloodstream and ailments of the lung;

Blue: Raises metabolism; is used to stabilize the heart, muscles and bloodstream; used to treat burns, skin diseases, glaucoma, measles and chicken pox, and throat problems;

Green: Strengthens bones and muscles, disinfects bacteria and virus, and relieves tension; used to treat . . . malaria, back problems, cancer, nervous disorders, and ulcers, and to manage heart problems and blood pressure. [2, p. 93]

The allocation of colors to each of the chakras is reminiscent of the doctrine of the four color-linked humors of the body from ancient Greek medicine: black bile, yellow bile, green phlegm,

and red blood. The linking of color with the humors, the four elements (earth, fire, water, and air) as well as the seasons, represented the color correspondence beliefs that emerged in antiquity and continued through to the Renaissance as evidenced by the color correspondences depicted in the engraving by Nicoletto Rosex [29].

Aside from ancient belief systems, much of the information currently available in popular media about color psychology echoes, if not directly quotes, the work of a number of key theorists, including Faber Birren, Robert Gerard, Kurt Goldstein, and Max Lüscher (e.g., see [1–3, 22, 25, 30–32]).

Birren, Gerard, Goldstein, and Lüscher

Extensively quoted by recent authors, Birren published more than 40 books and over 250 articles on color psychology, color therapy, as well as color application. Tunney advises that Birren was a leading authority on color in the mid-twentieth century who was retained as a color consultant by DuPont, General Electric, Sears, Roebuck and Company, and the United States Navy [33]. Birren championed an unambiguous, irrefutable, and universal causal link between color and human response, and he cites Goldstein’s assertion that “It is probably not a false statement if we say that a specific color stimulation is accompanied by a specific response pattern of the entire (human) organism” [34, p. 144]. Given his role as color consultant, it was in his interests to champion such a link as he was able to then provide advice addressing this link in a range of commercial contexts.

Goldstein, who published *The Organism* in 1939, was considered a significant authority on the psychological and physiological impacts of color by Birren, who quoted Goldstein in his *Color Psychology and Color Therapy* publication: “under the influence of red light, time is likely to be overestimated. Conversely, under the influence of green or blue light, time is likely to be underestimated” and “under red light, weights will be judged as heavier; under green light they will be judged lighter” [35, p. 211]. Birren as well Mahnke cites Goldstein’s 1942 study, which

discussed the perceived stimulating effects of red and the opposite effects of green [16].

Gerard, whose 1958 doctoral thesis (“Differential effects of colored lights on psychophysical functions”) was subsequently presented as a conference paper (“Color and emotional arousal”), is also frequently cited by Birren and Mahnke [15, 36]. Gerard’s key findings from his study on the arousal properties of red, blue, and white illumination, as reported by Wise, Wise, and Beach, include: “statistical differences between red-blue (illumination) conditions for all physiological measures except heart rate. . . responses to the white light varied but most often were similar to those of the red condition” [37, p. 5]. Wise et al. also note that Gerard, whose study involved a small sample of only 24 male university students, advised caution in terms of the generalizability of his findings. It is unfortunate that Gerard’s advice was not reported by subsequent authors.

Lüscher, who developed and published the Lüscher Color Test, included four basic colors (which he referred to as psychological primaries) in the color test: orange-red, bright yellow, blue-green, and dark-blue, as well as four auxiliary colors – violet, brown, black, and neutral gray [38]. Citing anecdotal evidence, Lüscher assigned specific associations and affective characteristics to each color. For example, (red) has a “stimulating effect on the nervous system – blood pressure increases, respiration rate and heartbeat both speed up”; while blue “has the reverse effect – blood pressure falls, heartbeat and breathing both slow down” [38, p. 12]. The Lüscher Color Test is essentially a ► [color preference](#) test and its use in personality testing and assessment has been soundly cautioned [39].

While the works of Birren, Gerard, Goldstein, and Lüscher have their place in the literature on color, it is important to note that their theories have, to a large extent, been debunked due to methodological shortcomings. In addition, more recent and methodologically more rigorous studies have superseded these earlier authors, and this, plus a number of other reasons discussed below, provides a compelling case for applying the principle of caveat emptor to information about color

psychology and color therapy found in mainstream media.

Color Psychology and Color Therapy: Caveat Emptor

Latin for “buyer beware,” the term caveat emptor suggests that in the absence of a warranty, the buyer is at risk, and therefore onus for carefully assessing goods and services prior to purchase remains with the buyer. O’Connor suggests there are a number of reasons why the principle of caveat emptor should prevail in regard to information about color psychology and color therapy [50].

Firstly, the existence of an irrefutable and universal causal link between color and an unlimited range of psychological, biological, and behavioral responses remains a largely unsupported hypothesis. While numerous studies exist that focus on the interface between color and human response, a systematic and critical review of over 200 studies was conducted for the National Aeronautics and Space Administration (NASA) in 1988. This study concluded that “there are no ‘hard-wired’ linkages between environmental colors and particular judgmental or emotional states” [37, p. 46].

In addition, while a number of studies exist which focus on the interface between color and human response, the findings of many of these studies are limited by the setting and context of the study, the size and composition of the sample group, as well as the limited range of colors used in the stimuli. Extrapolating beyond the data in such studies is methodologically unsound, especially when small sample sizes are used [40]. Very small sample groups were used by many early studies including the often-quoted studies of Goldstein and Gerard:

- Sample size of 3–5 [16]
- Sample size of 24 [15]
- Sample size of 48 [41]
- Sample size of 14 [42]
- Sample size of 25 [43]

In addition, many early studies focused on the color attribute of hue alone without any consideration of the attributes of tonal value or saturation, and many research studies used small colored

squares of cardboard as stimuli and the findings from such studies were extrapolated to a whole range of other contexts or situations. The findings of such studies have extremely limited relevance due to the methodological weakness of studying a complex and highly subjective phenomenon as color in isolation with limited stimuli.

Ancient wisdom is often cited as a reason to believe claims about color psychology and color therapy, with the underlying logic implying that ancient wisdom embedded within such claims represents evidential proof. For example, Kopacz cites the link between color and chakras of ancient Hindu scriptures as evidence, while Van Wagner suggests that chromotherapy was practiced by ancient cultures, including the Egyptians and Chinese [1, 2]. Such authors imply that not only is the wisdom of the ancients above question but it provides evidential proof of the veracity of various color-related claims. Clearly, ancient belief systems are not always a guarantee of veracity as evidenced by the superseded beliefs that the world is flat, and the sun and moon orbit the earth. Without disparaging the wisdom of the ancients, it is a fair comment that a proportion of ancient wisdom has been superseded by later scientific discovery. The existence of a link with ancient wisdom should not of itself be used as evidential proof of any claim regarding color psychology and color therapy.

Factoids are not facts, although many authors who write about color psychology and color therapy would have us believe otherwise. Coined by author Norman Mailer, factoids are suppositions or inventions presented as fact, and a comedic illustration is the line in the film *Anchorman*: "It's anchorman not anchor-lady, and that's a scientific fact!" [44]. Color psychology and color therapy information available in popular culture often appears as factoids rather than facts, especially when nil evidence is cited to support claims and assertions. Examples of unsubstantiated factoids include: "We are hard-wired to yellow as a stimulus. . . If you're environment is boring and time passes slowly, surround yourself with small amounts of yellow" and "People who dislike yellow often favour blue to calm themselves and feel

secure. If you drink coffee for a pick-me-up, try drinking it from a yellow cup" [23, p. 32]; similarly, "Yellow can also create feelings of frustration and anger. While it is considered a cheerful colour, people are more likely to lose their tempers in yellow rooms and babies tend to cry more in yellow rooms" [1].

The use of control groups, measures to control internal and external validity, and adequate levels of statistical significance are absent from some studies that focus on the interface between color and human response, and these impact negatively on the scientific rigor of such studies. It is empirically and methodologically unsound to draw generalized conclusions from such studies and transfer them to different settings and contexts [40, 45–47].

The fallacy of the single cause provides another reason to be cautious about claims regarding color psychology and color therapy. This fallacy suggests that one single cause for an outcome represents causal oversimplification, and under the post-positivist paradigm, recent theorists such as Hård and Sivik consider the interface between color and human response to be highly complex and open to the influence of a wide range of factors and mediating variables [48]. While correlation may indicate the existence of an association between one variable (such as color) and human response, correlation does not imply causation and suggesting a causal link without allowing for mediating variables is considered empirically and methodologically unsound [40, 45, 47].

Inherent bias, subjective validation, and data cherry picking in research studies always call into question the findings of such studies [40]. Specifically, "Scientists have a vested interest in promoting their work," and this is often the case for color theorists and researchers whose publications support their role as consultants [40, p. 337]. In addition, subjective validation is evident in many studies that focus on the interface between color and human response. This occurs when two unrelated events are judged to be related because of an expectation of such or because an existing hypothesis demands such a relationship. For

example, to apply the notion that “green relieves stress” to the following claims: “green silk wrapped around the shoulders can have a very therapeutic effect” and “Those who have a green work environment experience fewer stomach aches” indicates subjective validation and renders such claims invalid.

It is unwise to ignore any mediating variables that may impact a research study and this is particularly relevant to research that focuses on the interface between color and human response. In this context, mediating factors include an individual’s personality, personal bias and feelings, as well as cultural experiences and affective state. Mehrabian found that individual differences exist in terms of stimulus screening ability, and high screeners are able to automatically screen out less important components of environmental stimuli such as color and sound as opposed to low screeners [49]. Stimulus screening ability is rarely if ever mentioned in research studies that focus on color and human response nor is an individual’s own bias in terms of whether or not they consider color to have an impact on their responses to color.

To conclude, color psychology claims information found in mainstream media suggests that color prompts a range of different human responses: psychological, biological, and behavioral. Many of these claims lack substantiation in terms of empirical support, exhibit fundamental flaws (such as causal oversimplification and subjective validation), and are generally personal opinion and factoids masquerading as fact. Color psychology is a discipline that is not accepted within the medical profession and exists on the periphery of alternative medicine and therapies. It is therefore recommended that the principle of caveat emptor is applied when evaluating information about color psychology plus check with a medical practitioner or psychologist before acting upon such information.

Cross-References

- Color Preference
- Primary Colors

References

1. Cherry, K.: Color psychology: how colors impact moods, feelings and behaviours. About.com: Psychology. <http://psychology.about.com/od/sensationandperception/a/colorpsych.htm> (2014). Accessed 4 Jan 2014
2. Kopacz, J.: Color in Three-Dimensional Design. McGraw-Hill, New York (2003)
3. Logan-Clarke, V.: What is colour therapy? <http://www.colourtherapyhealing.com/> (2014). Accessed 4 Jan 2014
4. Brainard, G.C., Hanifin, J.P., Greeson, J.M., Byrne, B., Glickman, G., Gerner, E., et al.: Action spectrum for melatonin regulation in humans. Evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**(16), 6405–6412 (2001)
5. Harmatz, M.G., Well, A.D., Overtree, C.E., Kawamura, K.Y., Rosal, M., Ockene, I.S.: Seasonal variation of depression and other moods: a longitudinal approach. *J. Biol. Rhythm.* **15**(4), 344–350 (2000)
6. Kasper, S., Wehr, T.A., Bartko, J.J., Gaist, P.A., Rosenthal, N.E.: Epidemiological findings of seasonal changes in mood and behavior. *Arch. Gen. Psychiatry.* **46**, 823–833 (1989)
7. Stevens, R.G., Blask, D.E., Brainard, G.C., Hansen, J., Lockley, S.W., Provencio, I., et al.: Meeting report: the role of environmental lighting and circadian disruption in cancer and other diseases. *Environ. Health Perspect.* **115**(9), 1357–1362 (2007)
8. Skene, D.J., Lockley, S.W., Thapan, K., Arendt, J.: Effects of light on human circadian rhythms. *Reprod. Nutr. Dev.* **39**(3), 295–304 (1999)
9. Irlen, H.: Reading problems and Irlen coloured lenses. *Dyslexia Rev.* **8**(3), 4–7 (1997)
10. Lehl, S., Gerstmeier, K., Jacob, J.H., Frieling, H., Henkel, A.W., Meyer, R., et al.: Blue light improves cognitive performance. *J. Neural Transm.* **114**, 457–460 (2007)
11. Warman, V.L., Dijk, D.J., Warman, G.R., Arendt, J., Skene, D.J.: Phase advancing human circadian rhythms with short wavelength light. *Neurosci. Lett.* **342**, 37–40 (2003)
12. Webb, A.: Considerations for lighting in the built environment: non-visual effects of light. *Energ. Buildings.* **38**, 721–727 (2006)
13. Mikellides, B.: Emotional and behavioural reactions to colour. In: Sivik, L. (ed.) *Colour and Psychology: from AIC Interim Meeting 96 in Gothenburg. Colour Report F50.* Scandinavian Colour Institute, Stockholm (1997)
14. Ali, M.R.: Pattern of EEG recovery under photic stimulation by light of different colors. *Electroencephalogr. Clin. Neurophysiol.* **33**, 332–335 (1972)
15. Gerard, R.M.: Differential effects of colored lights on psychophysiological functions. University of California, Los Angeles, Unpublished Ph.D. thesis (1958)
16. Goldstein, K.: Some experimental observations concerning the influence of colors on the function of

- the organism. *Occup. Ther. Rehabil.* **21**(3), 147–151 (1942)
17. Mikellides, B.: Color and psychological arousal. *J. Archit. Plann. Res.* **7**(1), 13–20 (1990)
 18. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color. Res. Appl.* **32**(2), 144–150 (2007)
 19. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modelling. *Color. Res. Appl.* **29**(5), 381–389 (2004)
 20. Pantone: color psychology. <https://store.pantone.com/uk/en/colour-psychology> (2020). Accessed 19 Oct 2020
 21. Campbell, L.: Mind body healing – how it works. <https://consciouslivingmagazine.com.au/> (2009). Accessed 19 Oct 2020
 22. Wright, A.: Psychological properties of colours. <http://www.colour-affects.co.uk/how-it-works> (2014). Accessed 11 Jan 2014
 23. Rewell, C.: What do colours mean? Part II. <https://www.wellbeing.com.au/mind-spirit/mind/what-do-colours-mean-part-ii.html> (2009). Accessed 19 Oct 2020
 24. Hill, T.R.: Using Color to Create Healing Environments (Report Prepared for DuPont Corian). Little Fish Think Tank/DuPont, Atlanta (2008)
 25. Mahnke, F.: Color, Environment and Human Response. Wiley, New York (1996)
 26. CTH: Colour therapy healing workshop. <http://www.colourtherapyhealing.com> (2009). Accessed 10 Apr 2009
 27. ISCD: International school of colour and design. <http://www.iscd.edu.au/default2.asp> (2009). Accessed 10 Apr 2009
 28. NCC: Nature care college. <http://naturecare.com.au> (2009). Accessed 10 Apr 2009
 29. Gage, J.: Colour and Culture. Thames & Hudson, London (1995)
 30. Feisner, E.A.: Colour: how to Use Colour in Art and Design. Laurence King, London (2000)
 31. Holtzschue, L.: Understanding Color. Wiley, New York (2006)
 32. Stone, T.L., Adams, S., Morioka, N.: Color Design Workbook. Rockport, Beverly (2006)
 33. Tunney, S.: Guide to the Faber Birren papers. <https://archives.yale.edu/repositories/5/resources/364> (1998). Accessed 19 Oct 2020
 34. Birren, F.: Color Psychology and Color Therapy. Citadel, Secaucus (1961)
 35. Goldstein, K.: The Organism: a Holistic Approach to Biology Derived from Pathological Data in Man. Zone, New York (1995 [1939])
 36. Gerard, R.M.: Color and emotional arousal (abstract from the program of the 66th annual convention of the American Psychological Association). *Am. Psychol.* **13**(7), 340 (1958)
 37. Wise, B.K., Wise, J.A., Beach, L.R.: The human factors of color in environmental design: a critical review. NASA Grant no. NCC 2–404. NASA Ames research Centre. Moffett Field. (1988)
 38. Lüscher, M.: The Lüscher Color Test. Random House, New York (1969)
 39. Walters, J., Apter, M.J., Svebak, S.: Color preference, arousal, and the theory of psychological reversals. *Motiv. Emot.* **6**(3), 193–215 (1982)
 40. Sutherland, W.J., Spiegelhalter, D., Burgman, M.A.: Twenty tips for interpreting scientific claims. *Nature.* **503**, 335–337 (2013)
 41. Nakshian, J.S.: The effect of red and green surrounding on behavior. *J. Gen. Psychol.* **70**(1), 143–161 (1964)
 42. Nourse, E.W., Welch, R.B.: Emotional attributes of color: a comparison of violet and green. *Percept. Mot. Skills.* **32**(2), 403–406 (1971)
 43. Goodfellow, R.A., Smith, P.C.: Effects of environmental color on two psychomotor tasks. *Percept. Mot. Skills.* **37**(1), 296–298 (1973)
 44. Ferrell, W., McKay, A., Apatow, J., McKay, A.: Anchorman: the Legend of Ron Burgundy. DreamWorks Pictures, Los Angeles (2004)
 45. Argyrous, G.: Statistics for Social and Health Research. Sage, London (2001)
 46. Campbell, D.T., Stanley, J.C.: Experimental and Quasi-Experimental Designs for Research. Houghton Mifflin, Boston (1963)
 47. Coolican, H.: Research Methods and Statistics in Psychology, 4th edn. Hodder & Stoughton, London (2004)
 48. Hård, A., Sivik, L.: A theory of colors in combination – a descriptive model related to the NCS color-order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
 49. Mehrabian, A.: Individual differences in stimulus screening and arousability. *J. Pers.* **45**, 237–250 (1977)
 50. O'Connor, Z.: Colour psychology and colour therapy: Caveat emptor. *Color Res. Appl.* **36**(3), 229–234 (2011)

Color Qualia

► Color Phenomenology

Color Range

► Palette

Color Roles

► Functionality of Color

Color Scene Statistics, Chromatic Scene Statistics

Yoko Mizokami
Graduate School of Engineering, Chiba
University, Chiba, Japan

Synonyms

Chromatic image statistics; Color environment;
Color image statistics; Color statistics; Natural
color distribution; Natural scene statistics

Definition

Color scene statistics or chromatic scene statistics are statistical characteristics of scene color. There are a number of ways to analyze color scenes (natural scenes) statistically in relation to vision, such as the average color, color distribution, Bayesian model, principal component analysis, and probabilistic models. Color scene statistics are closely related to the evolution and development of the color vision mechanism at every level of the visual system.

Overview

Color vision has evolved with the natural environment, including the natural illumination from the sun and the sky. Thus, the color statistics of a natural scene must have an enormous impact on the evolution and development of the color vision mechanism. There have been many attempts to analyze natural scene statistics and find the connection to the color vision mechanism as well as other aspects of the visual mechanism.

Traditionally, the average color and color distribution of a scene are associated with color adaptation and color constancy. Based on those statistical data, physiological, empirical, and statistical models of color constancy were introduced. Other aspects of visual perception, such as color discrimination or color appearance, would be formed by the environment to obtain the best performance or representation under a given color environment. Measurement techniques, such as two- and three-dimensional measurements of scenes and multispectral measurement, provide additional information for deeper analysis or analysis from different perspectives. Powerful computational analyses introduce new statistical approaches for investigating the mechanism of color vision. Here, the topics of color scene statistics related to color vision, visual perception, and their evolution are discussed.

Measurement of Objects, Illumination, and Scenes

Obtaining accurate color information of scenes is important to analyze color scene statistics. Camera calibration techniques and calibrated images in which the RGB values of images are transferable to information on chromaticity coordinates have been developed.

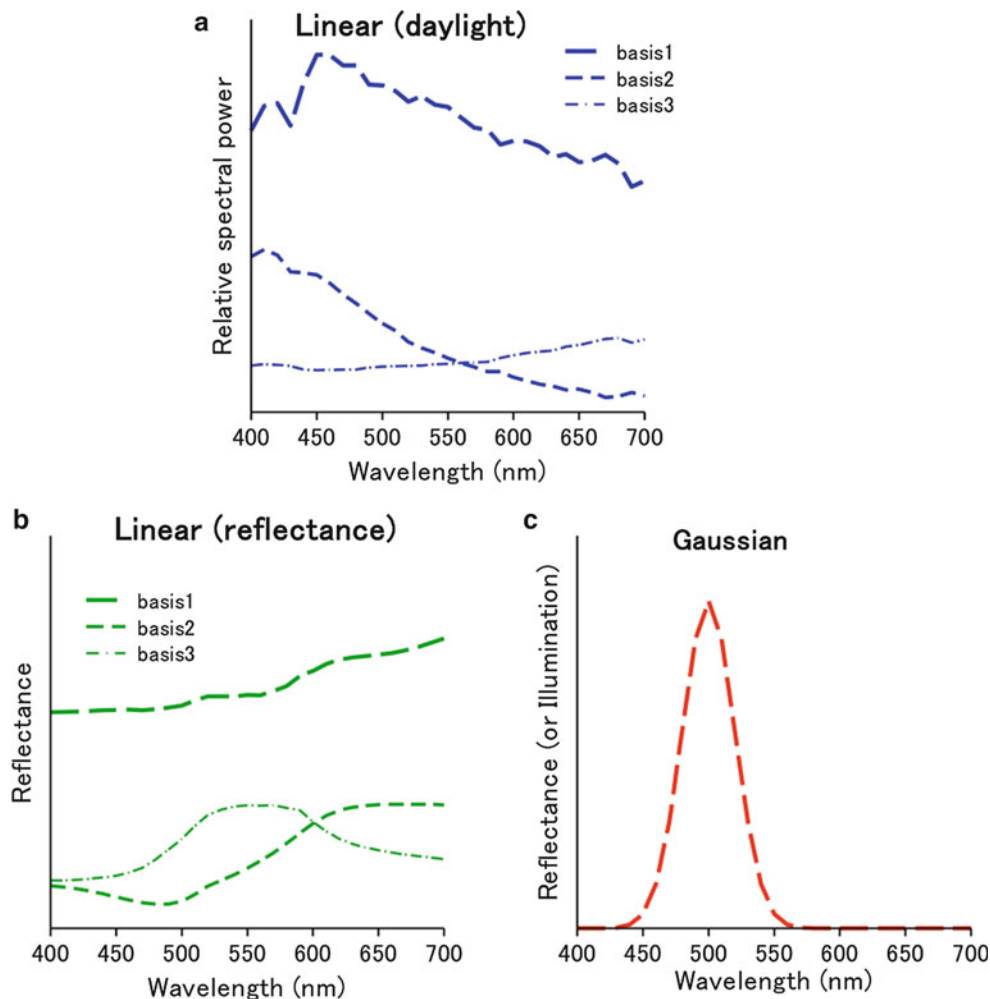
Another important aspect of analyzing colors in natural scenes is spectral information. RGB information has only three dimensions, and consequently, some information is lost. For instance, RGB information cannot differentiate metameric color, which results when a pair of stimuli exhibit the same apparent color but with different spectral power distributions. Many groups have measured the daylight spectra and spectral reflectance of natural and printing surfaces. Published databases (e.g., SOCS [1]) and personal databases on websites are available.

For obtaining the spectral color information of an entire scene, multispectral or hyperspectral image capture techniques using a multispectral camera are used. Those data are used for studies on color vision mechanisms. It is difficult, however, to obtain a hyperspectral image manually because it

takes a long time, and the scene must be still (no moving objects or wind) to capture an image with multiple wavelengths by changing band-pass filters. Therefore, collections of multispectral images of outdoor scenes are limited [2]. The recent development of fast and accurate multispectral cameras will allow further statistical analysis based on a large number of multispectral data sets.

Although multispectral information is important and useful, it includes a massive amount of data that is inconvenient to handle. Thus, there have been many attempts to represent multi-dimensional data with a smaller number of

dimensions by compressing the amount of information using principal component analysis (PCA). The natural spectra can be represented well by three basis functions. Those for the illuminant obtained by Judd [3] are shown in Fig. 1. The basis functions give the relative spectral radiant power distribution of the CIE daylight illuminant at different color temperatures (Fig. 1a). Cohen [4] applied this analysis to the reflectance of Munsell color paper and showed three basis functions that were different from those obtained by Judd. Later studies suggested that Cohen's data were applicable to natural surfaces (see Fig. 1b)



Color Scene Statistics, Chromatic Scene Statistics, Fig. 1 Examples of three basis functions for daylight (a) and reflectance (b), respectively. A Gaussian fitting function

with three parameters: peak, amplitude, and bandwidth (for both illumination and reflectance) is also shown (c)

and that three basis functions were necessary and probably sufficient for representing the spectral reflectance functions of natural objects. (Note that some studies suggested that a larger number of basis functions are needed depending on the context.) These results led to the hypothesis that the human visual mechanism has a similar representation in the visual process [5]. However, it was recently suggested that the Gaussian fitting function with three parameters: peak, amplitude, and bandwidth to natural spectra (see Fig. 1c) is as good as the fitting by three basis functions [6]. More investigations would be needed to determine how the human visual system represents the spectral world.

Analysis of Natural Color Statistics

There are a number of ways to analyze the color properties of a scene statistically. The most basic method would be by average color and color distribution. Analysis taking into account spatial frequency is also important because the spatial contrast sensitivities of luminance and the L-M (reddish–greenish) and S-(L + M) (bluish–yellowish) opponent-color channels are different. Those chromatic channels tend to have a low-pass shape, suggesting the low-frequency component contributes more to color perception. Recent developments in statistical modeling, along with powerful computational tools, have enabled researchers to study more sophisticated statistical models of visual images and to evaluate these models empirically against large data sets [7].

Based on the information obtained from color scene statistics, the mechanisms of vision and color perception have been investigated from aspects of the peripheral to central visual system to perspectives on evolution, development, and short- to long-term adaptation, as described in the following sections.

Effect on Evolution of Color Vision

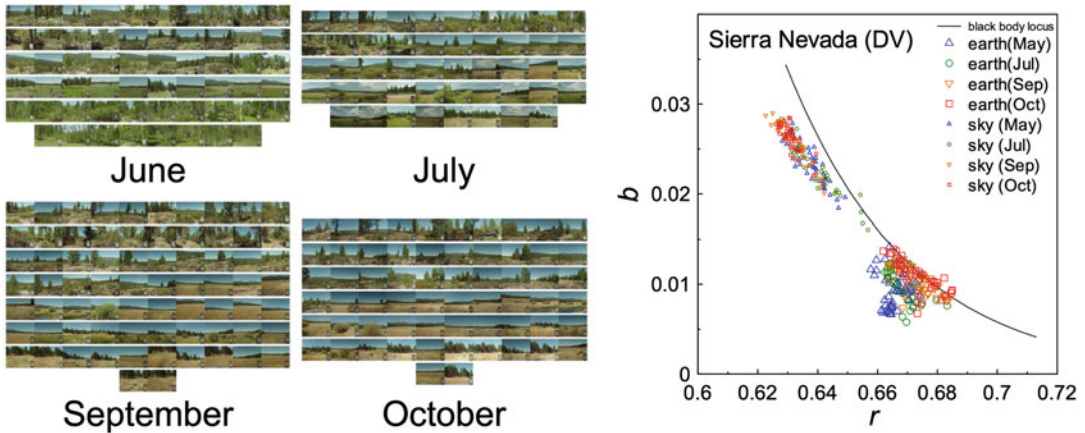
The sensitivity of the photoreceptors and how the cone signals are combined in postreceptoral

channels have been explained by assuming that they optimize the efficient coding of natural color signals and that they are tuned to specific signals.

For example, it has been suggested that trichromacy in Old World primates and the reflectance functions of tropical fruits coevolved and that primate color vision has been shaped by the need to find reddish (ripe) fruit or young (edible) leaves among green foliage, as shown by the example in Fig. 2. For instance, the tuning of the L and M cones and the later opponent-color system (L-M) based on the signal representing their difference are optimized to detect the color signals provided by ripening fruit or edible foliage [8]. This hypothesis is a good example of how color statistics of the visual environment



Color Scene Statistics, Chromatic Scene Statistics, Fig. 2 Examples of the advantage of trichromacy (*top*), capable of red–green discrimination, compared to dichromacy (*bottom*) on finding red fruits among green leaves



Color Scene Statistics, Chromatic Scene Statistics, Fig. 3 An example of seasonal changes in the natural color scene (Mountain forest in the Sierra Nevada) and the average chromaticities for earth (large symbols) and

sky (small symbols) for individual scenes plotted in the Macleod-Boynton color space. b axis represents $S/(L + M)$, r axis represents the $L/(L + M)$

affected the early-stage evolution of the color vision system.

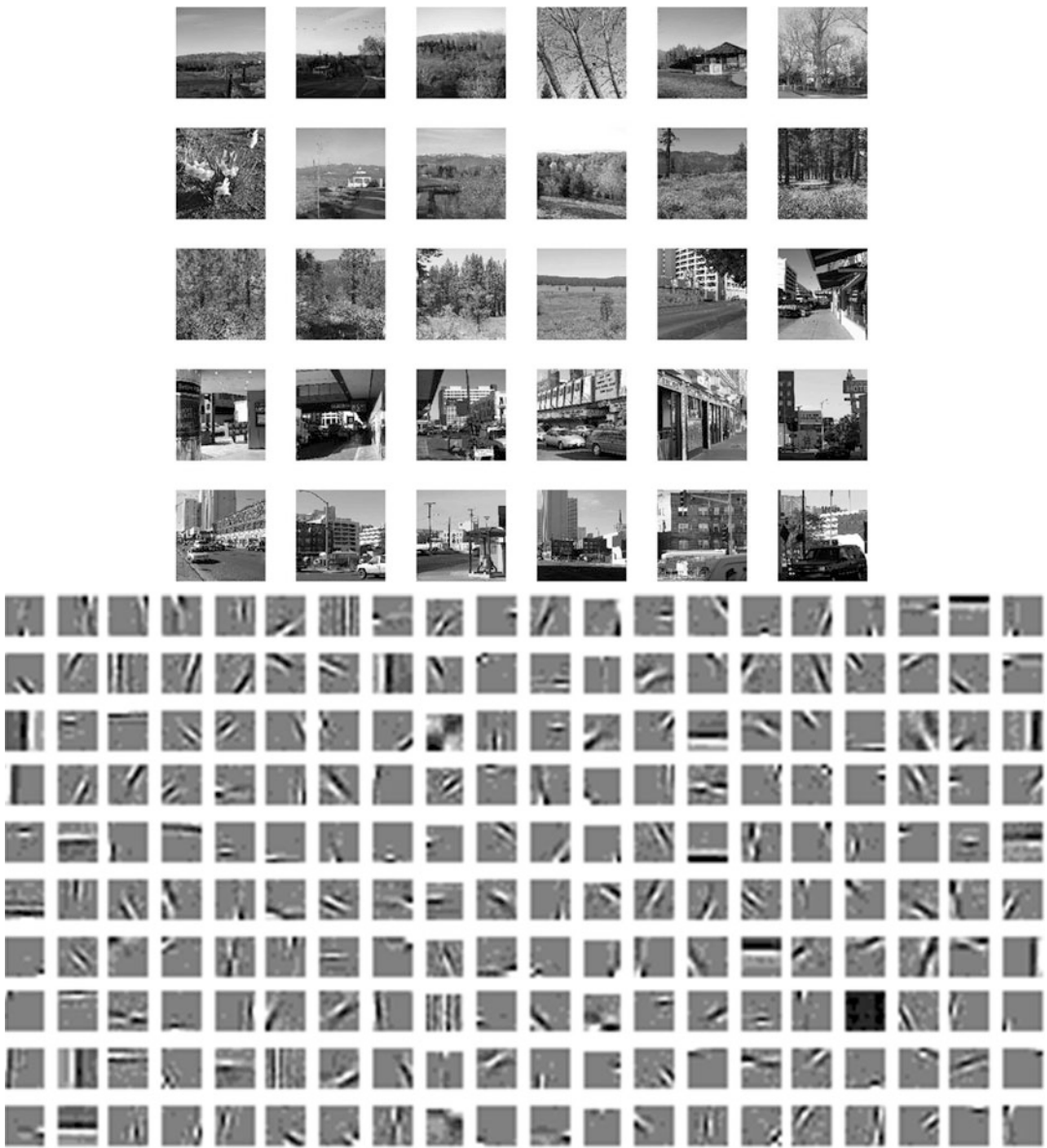
It has also been shown that seasonal variations in the color statistics of natural images alter the average color in scenes, as shown in Fig. 3. The average chromaticities for earth and sky for individual scenes in the Macleod-Boynton color space indicate that large seasonal color shift along b (+ L) direction. Color distribution in scenes also changes across seasons. On opponent-color space, arid periods are marked by a mean shift toward the + L pole of the L - M chromatic axis. A rotation in the color distributions away from the S -($L + M$) chromatic axis and toward an axis of bluish–yellowish variation, both primarily due to changes in vegetation. These data imply that systematic color changes in season contribute to the construction of the mechanism of both visual sensitivity and color appearance [9].

There are other hypotheses, such as skin tone discrimination and predator detection affecting evolution. There are a number of variations or changes in natural environments other than the examples shown above, which may contribute to the formation of the color vision mechanism. It is interesting to pursue how the color vision mechanism was optimized during evolution and development [10].

Cortical Mechanism and Probabilistic Approaches

Cortical color coding is also thought to have evolved to represent important characteristics of the structure of color in the environment. Probabilistic modeling allows us to test the efficient coding hypothesis for both individual neurons and populations of neurons experimentally [7].

“Sparse coding,” the concept that neurons encode sensory information using a small number of active neurons at any given second, has been introduced. It increases the storage capacity of associative memories and saves energy. It also makes the structure of natural signals explicit and makes complex data easier to interpret during visual processing [11]. They created a model based on a linear superposition of basis functions. They adapted these functions to maximize the sparsity of the representation while preserving information in the images by maintaining a bound on the mean squared reconstruction error. Function sets that emerge after training on numerous image patches randomly extracted from natural scenes resemble the spatial receptive field properties of simple cells (i.e., spatially localized, oriented, and band-pass in different spatial frequencies). Figure 4 shows an example of a set of basis functions



Color Scene Statistics, Chromatic Scene Statistics, Fig. 4 An example of sparse coding created by a procedure by the method in the paper of Olshausen and Field [11]. *Top*, image-set for learning; *Bottom*, derived basis functions

derived from a set of learning images, including natural and man-made scenes, based on a sparse coding model. Although this is only a rough example based on the limited number of images, it shows that any natural image (scene) can be constructed using combinations of basis functions. This suggests that the human visual system also constructs the image of a natural scene based on the

activity of neurons having receptive fields with different properties.

Probabilistic models are also applied to the chromatic structure of natural scenes. Independent component analysis (ICA) has been applied not only to gray images but to natural color images to establish an efficient representation of color in natural scenes. It was shown to produce

achromatic and color-opponent basis functions as well as spatiochromatic independent components that resemble cortical neurons [12]. This suggests a relationship between statistics of natural scenes and cortical color processing.

Effect of Color Statistics of a Scene on Color Constancy

Color constancy is a phenomenon in which the stability of the color appearance of an object's surface is maintained among variations in the illuminating environment. The central issue is always how humans separate illumination and the surface of objects, which is a necessary task for achieving color constancy. Identical to other mechanisms in earlier stages of the visual system, color constancy is also associated with the color statistics of environments [5, 10, 13].

A classic approach is the von Kries adaptation. To set a neutral point (or illumination color) for applying the von Kries adaptation, many assumptions or hypotheses based on scene statistics have been proposed (e.g., the average color of a scene: the gray world assumption). In Bayesian decision theory, which is a more statistical approach, the most likely combination of illuminants and surfaces is determined based on prior distributions that describe the probability that particular illuminants and surfaces exist in the world.

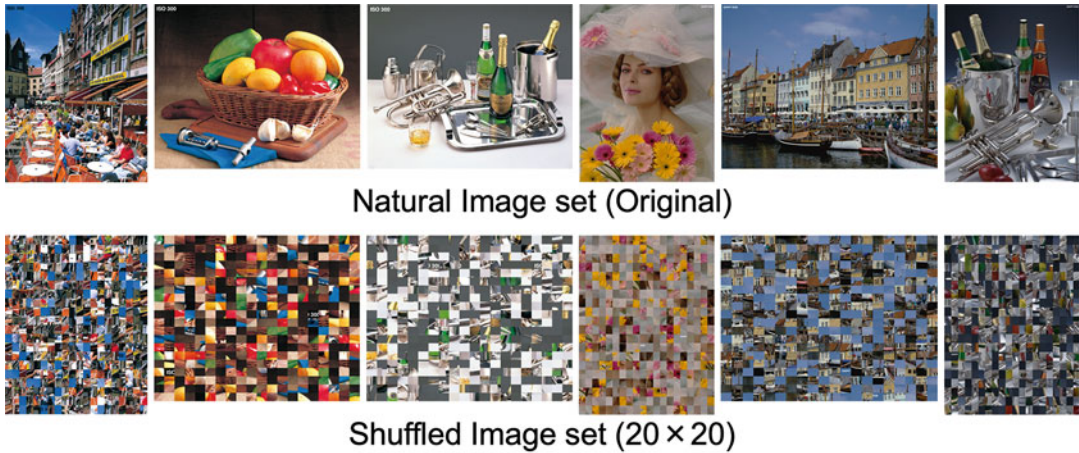
Characteristics of scene statistics also contribute to color constancy [13, 14]. One question is how humans can tell the difference between a white paper under red illumination and a red paper under white illumination if the papers have the same chromaticity. It was suggested that the luminance of the white paper decreases statistically under the red light and that this relationship could be a clue for differentiating the two. Besides, in natural scenes, chromatic variations and the luminance variations aligned with them mainly arise from object surfaces, such as the border between different materials. In contrast, pure or near-pure luminance variations mainly arise from inhomogeneous illumination, such as shadows or shading. The human visual system uses these color–luminance relationships and

determines the three-dimensional structure of a scene from the natural relationships between color and luminance in the visual scene, material surface, and so on. It is suggested that natural scene statistics could also be the universal basis of color context effects, such as color contrast and color assimilation.

Adaptation to the Color of Natural Scenes

As shown in color constancy, it is well known that color perception adapts to the hue change of color distribution. Color appearance is also influenced by the saturation or variance of the color distribution of a scene. A pale color patch appears less colorful when surrounded by saturated colors, and chromatically selective compression along any direction of chromatic variation occurs after adapting to temporal color variation. These suggest adaptation to the specific color gamut within individual scenes and natural environments [10, 13]. It is also shown that the impression of a natural image shifts to being less colorful after adaptation to a series of saturated images and vice versa [15]. Moreover, it is suggested that simple statistics, such as the color distribution itself, cannot explain the effect because the adaptation is stronger for natural images than shuffled images, as shown in Fig. 5. It implies the contribution of higher-level mechanisms, such as scene recognition or cognition.

The timescale of adaptation varies. It can be seconds, minutes, hours, months, years, and possibly a lifetime. The tuning of sensitivities of the color vision mechanism likely continues in all timescales. The evolution and development of the visual function could be considered a very long-term adaptation. It is also suggested that more cognitive or social aspects of color perception, such as color category, could be influenced by the color scene statistics of each region. One of the functional meanings of adaptation would be compensating for variations within an observer and between observers, as well as within an environment and between environments, to maintain a stable and coherent color appearance. The color



Color Scene Statistics, Chromatic Scene Statistics, Fig. 5 An example of natural and shuffled images used for the colorfulness adaptation experiment [15]

scene statistics certainly contribute to it. The adaptation shaped by environmental pressure would achieve an efficient transmission and stability of information from the periphery of the visual system to the centers of the brain.

Cross-References

- [Appearance](#)
- [Color Constancy](#)
- [Color Contrast Adaptation](#)
- [Environmental Influences on Color Vision](#)
- [Evolution of Primate Color Vision](#)
- [Hyperspectral Imaging](#)
- [Spectral Power Distribution](#)

References

1. Standard object colour spectra database for colour reproduction evaluation (SOCS). TR X 0012:2004 (ISO/TR 16066:2003), Japanese Standards Association (2004)
2. Foster, D.H., Nascimento, S.M., Amano, K.: Information limits on neural identification of colored surfaces in natural scenes. *Vis. Neurosci.* **21**, 331–336 (2004). <https://doi.org/10.1017/S0952523804213335>
3. Wyszecki, G., Stiles, W.S.: *Color Science*. Wiley, New York (1982)
4. Cohen, J.: Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci.* **1**, 369–370 (1964). <https://doi.org/10.3758/BF03342963>
5. Foster, D.H.: Color constancy. *Vis. Res.* **51**, 674–700 (2011). <https://doi.org/10.1016/j.visres.2010.09.006>
6. Mizokami, Y., Webster, M.A.: Are Gaussian spectra a viable perceptual assumption in color appearance? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A10–A18 (2012). <https://doi.org/10.1364/JOSAA.29.000A10>
7. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001). <https://www.annualreviews.org/doi/10.1146/annurev.neuro.24.1.1193>
8. Osorio, D., Vorobyev, M.: A review of the evolution of animal colour vision and visual communication signals. *Vis. Res.* **48**, 2042–2051 (2008). <https://doi.org/10.1016/j.visres.2008.06.018>
9. Webster, M.A., Mizokami, Y., Webster, S.M.: Seasonal variations in the color statistics of natural images. *Network.* **18**, 213–233 (2007). <https://doi.org/10.1080/09548980701654405>
10. Webster, M.A.: Adaptation and visual coding. *J. Vis.* **11**(5), 1–23 (2011). <https://doi.org/10.1167/11.5.3>
11. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature.* **381**, 607–609 (1996). <https://doi.org/10.1038/381607a0>
12. Hyvärinen, A., Hurri, J., Hoyer, P.O.: *Natural Image Statistics: a Probabilistic Approach to Early Computational Vision*. Springer, London (2009)
13. MacLeod, D.I.A., Golz, J.: A computational analysis of colour constancy. In: Mausfeld, R., Heyer, D. (eds.) *Colour Perception: Mind and the Physical World*, pp. 205–246. Oxford University Press, London (2003)

14. Shevell, S.K., Kingdom, F.A.A.: Color in complex scenes. *Annu. Rev. Psychol.* **59**, 143–166 (2008). <https://doi.org/10.1146/annurev.psych.59.103006.093619>
15. Mizokami, Y., Kamesaki, C., Ito, N., Sakaibara, S., Yaguchi, H.: Effect of spatial structure on colorfulness adaptation for natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A118–A127 (2012). <https://doi.org/10.1364/JOSAA.29.00A118>

Color Scheme

Nilgün Olguntürk

Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

Synonyms

[Color combination](#); [Color harmony](#); [Color palette](#)

Definition

An organized selection and arrangement of colors in design.

Introduction and Classification of Color Schemes

Colors from the color circle are usually combined in a systematic and logical manner to create a color scheme. The number of colors (different hues, saturations and lightnesses) in a scheme could range from two to many. A very basic color scheme would be to use black text on a white background, a common default color scheme in Web design.

A predetermined design idea about the final appearance of two-dimensional designs or products or the ambiance or atmosphere in interiors usually shapes the systematic or logic behind combining different colors into establishing a

color scheme. Creating style and appeal, evoking intended feelings (such as in an ambiance), and establishing more practical outcomes all are concerns while deciding on a color scheme.

Different types of schemes are used. These are predominantly based on the selection of colors that are regarded as being harmonious, in other words are aesthetically pleasing when viewed together. This can either be achieved with similarities or with contrasts [1, 2]. As color has three dimensions, namely, hue, saturation, and lightness, color schemes might use colors with similar hues, saturations, or lightnesses, as well as colors with contrasting hues, saturations, or lightnesses.

Color Schemes Using Harmony of Similarities

These schemes use similarity of hues (namely, monochromatic and analogous color schemes), similarity of saturations, or similarity of lightnesses. Monochromatic color schemes use a single hue and obtain a variety of colors with variations in that single hue's saturation and/or lightness levels. For example, in a room, a monochromatic color scheme may use only a certain blue with all its varying lightnesses and darkneses (lightness level) or its paler and vivid (saturation level) versions. Analogous color schemes use neighboring hues in a color circle [3]. The hues chosen in this scheme may incorporate secondary hues (hues that can be produced by adding two primary hues) such as using green, yellow-green, and yellow together. There are also warm or cool color schemes. These schemes either use all warm hues (e.g., red, orange, yellow) or all cool hues (e.g., blue and green) together.

Schemes using similarity of lightnesses combine different hues with similar lightness levels. For instance, they use only light colors together or only dark colors together. Similarities of saturations might also be used where only weak colors or only pure, vivid colors are combined together. In achromatic color schemes, only achromatic colors (black, white, and grays) are used. In this case if the design or space becomes monotonous, sometimes an accent color is added to these schemes, such as a red or a yellow.

Color Schemes Using Harmony of Contrasts

These schemes use contrasts of hues, contrasts of saturations, or contrasts of lightnesses [4]. Schemes with contrasting hues use opposite hues on the color circle. There are different complementary contrast color schemes depending on the desired number of hues to be used, namely, complementary contrast color schemes (two hues), split-complementary color schemes (three hues), double-complementary color schemes (four hues), etc. Complementary contrast color schemes use two opposing hues on the color wheel, for example, purple and yellow or red and green. These two hues can be used in the design with their many varying lightnesses or saturations. The name complementary contrast derives from afterimage complementaries, where the human visual system tries to compensate the overexposure of a certain hue with its opposing hue in the visual system [5]. When one looks at red for a few minutes, where red is the one and only thing in that person's visual field, and then that person looks at a white surface, that person will see an afterimage of the opposing hue (a light green) on that white surface. Split-complementary color schemes use not the direct opposite hue of a selected hue, but its immediate two neighbors. For example, blue will be used not with its exact opposite orange, but its two immediate neighbors which are yellow-orange and red-orange. Double-complementary color schemes use four hues, mainly two hues next to each other and their direct opposites on the color circle. Thus, yellow and yellow-orange are used together with purple and purple-blue. Triadic color schemes use hues that are evenly spaced around the color wheel, thus having equal number of hues in between them, such as using green, orange, and purple in a design.

Schemes using contrast of lightnesses use light and dark colors together, and schemes using contrast of saturations use weak and strong colors together. It is also possible to create a color scheme by using achromatic and chromatic colors together.

As there are many colors that human beings can see which can come together in many different combinations as described earlier, the initial

intent of the design or the design idea becomes prominent before deciding on the color scheme. The color scheme is chosen usually to best reflect the design intent or the design idea. This design intent or idea may vary from being fully functional to purely aesthetic. Some common design intentions are evoking an emotion, hiding from vision (camouflage), making something visible, or coding. If the design intent is to evoke an emotion, for example, creating a calming ambience, and the design idea is connotations with nature, the color scheme for a place might use similarity of lightnesses and saturations, with hues such as brown, green, and blue. If the design intent is to make things visible, for example, for visually impaired people, then the designer should not use schemes with similar lightnesses. In this case, using dark blue (navy), dark red, and black together would not make the design visible for visually impaired or elderly people.

A color scheme is a selection and arrangement of colors in an organized manner. The selection of a certain color scheme over another one purely depends on the design intent and the design idea of a product or space.

Cross-References

- [Color Circle](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Palette](#)
- [Primary Colors](#)

References

1. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors*. Schiffer, West Chester (1987)
2. Birren, F.: *Principles of Color*. Van Nostrand Reinhold, West Chester (1987)
3. Itten, J.: *The Elements of Color*. In: Birren, F. (ed.). (trans: Van Hagen, E.). Van Nostrand Reinhold, New York (1970)
4. Itten, J.: *Design and Form*. Thames and Hudson, London (1987)
5. Kuehni, R.G.: *Color: Essence and Logic*. Van Nostrand Reinhold, Berkshire (1983)

Color Science

- ▶ [Hyperspectral Imaging](#)
-

Color Selection

- ▶ [Palette](#)
-

Color Signals

- ▶ [Environmental Influences on Color Vision](#)
-

Color Similarity Space

- ▶ [Psychological Color Space and Color Terms](#)
-

Color Solids

- ▶ [Color Order Systems](#)
-

Color Sorting

- ▶ [Color Constancy and Categories](#)
-

Color Spaces

- ▶ [Color Order Systems](#)
-

Color Spreading

- ▶ [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Color Spreading, Neon Color Spreading, and Watercolor Illusion

Baingio Pinna

Department of Biomedical Sciences, University of Sassari, Sassari, Italy

Synonyms

[Color spreading](#); [Filling-in](#); [Visual illusions](#)

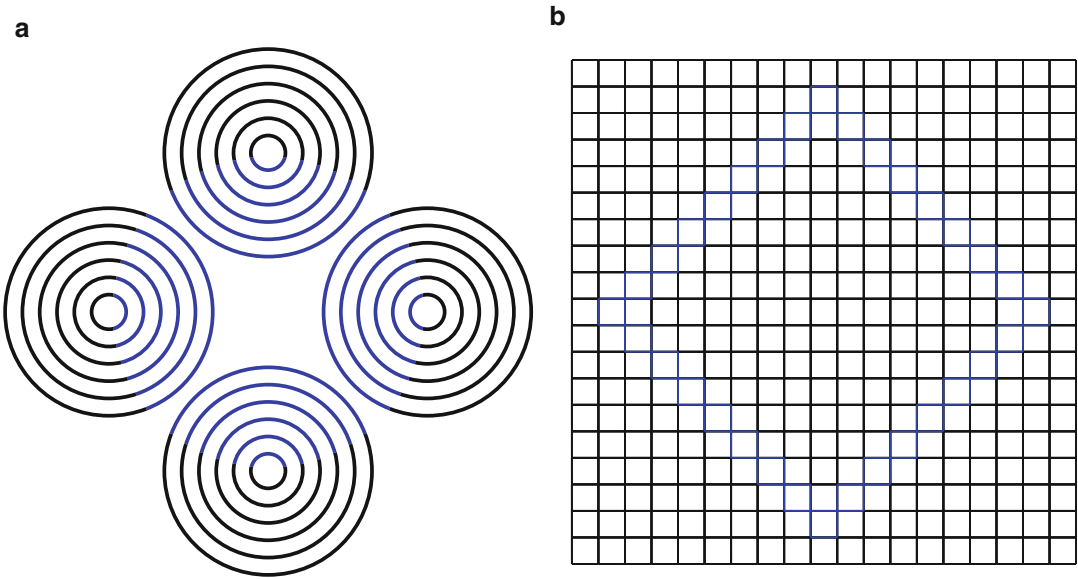
Definition

The color spreading is a long-range assimilative spread of color emanating from a thin-colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. The two main examples of color spreading are the well-known neon color spreading and the watercolor illusion.

Neon Color Spreading

In 1971, Varin [1] studied a novel “chromatic spreading” phenomenon induced when four sets of concentric black circumferences, arranged in a cross-like shape, are partially composed of blue arcs, thus producing a virtual large central blue circle (Fig. 1a). The virtual circle appears as a ghostly transparent circular veil of chromatic translucent diffusion of bluish tint spreading among the boundaries of the blue arcs and filling in the entire illusory circle, induced by the terminations of the black arcs (Fig. 1).

This phenomenon was independently reported a few years later from Varin’s discovery by van Tuijl [2], who named it “neon-like color spreading.” He used a lattice of horizontal and vertical black lines, where segments of different colors (e.g., blue) create an inset diamond shape. The main outcome reveals again a tinted transparent diamond-like veil above the lattice (Fig. 1b).



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 1 Neon color spreading: The central virtual circle (a) and the inset virtual diamond shape (b)

Geometrically, the main geometrical property of all the known cases of neon color spreading is the continuation of one contour, usually black, in a second contour with a different color or, differently stated, a single continuous contour changing from one color to another. Phenomenally, color spreading manifests coloration and a figural effect described in detail in the following sections.

Coloration Effects in Neon Color Spreading

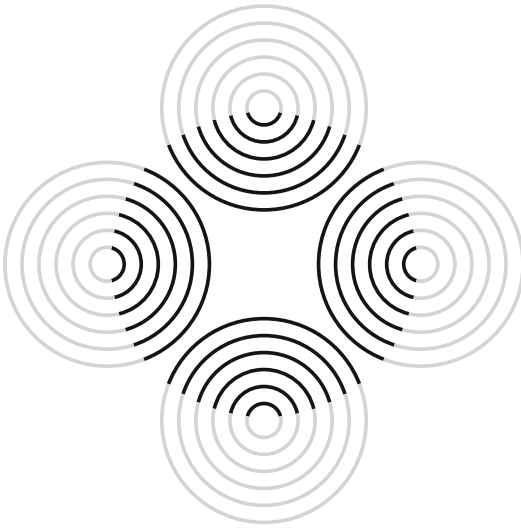
The phenomenology of the coloration effect within neon color spreading depends on the luminance contrast between the two inducing contours and is summed in the next points [2, 3]. (i) The color is perceived as a diffusion of a certain amount of pigment of the inset chromatic segments. (ii) The appearance of the coloration is diaphanous like a smoggy neon upon the background or (under achromatic conditions) like a shadowy, dirty, or filmy transparent veil. (iii) Under conditions where the inset figure is

appear as a ghostly overlapping transparent veil of bluish tint spreading among the boundaries of the blue components

achromatic and the surrounding inducing elements chromatic, the illusory veil of the inset figure appears tinted not in the achromatic color of the embedded elements, but in the complementary color of the surrounding elements, e.g., the achromatic components appear to spread greenish or bluish illusory colors, respectively, with red or yellow inducers.

Figural Effects in Neon Color Spreading

The apparent coloration of neon color spreading is related to its figural effects. Phenomenally, (i) the illusory neon coloration manifests a depth stratification appearing in front of the inducing elements; (ii) it is also perceived as a transparent film; (iii) by reversing the relative contrast of inset versus surrounding components, the depth stratification reverses accordingly (Fig. 2); (iv) the illusory colored region, under different chromatic conditions, may assume different figural roles by becoming, for example, a superimposed “light,” a “veil,” a “shadow,” or a “fog.”

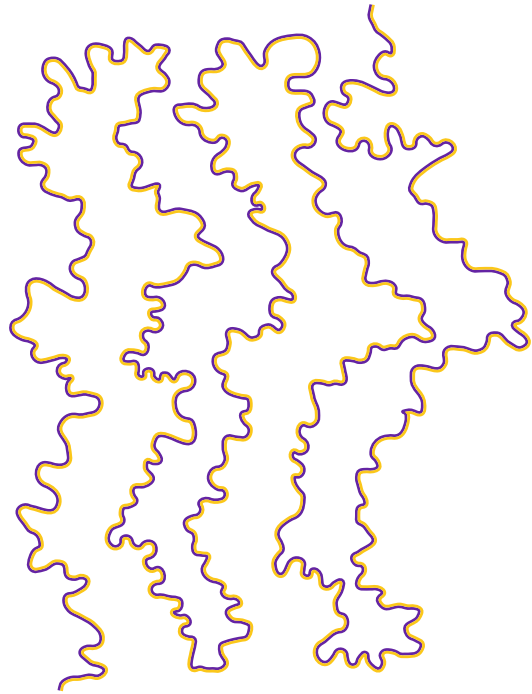


Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 2 By reversing the relative contrast of inset versus surrounding components (cf. Fig. 1), the depth stratification of neon color spreading reverses accordingly

Watercolor Illusion

The “watercolor illusion” is a long-range spread of color diffusing from a thin-colored contour adjacent to a darker chromatic contour and imparting a clear figural effect within large regions [4–9]. In Fig. 3, purple wiggly contours flanked by orange edges are perceived as undefined curved solid shapes, similar to peninsulas emerging from the bottom, evenly colored by a light veil of orange tint spreading from the orange edges.

By reversing purple and orange contours of Fig. 2, the coloration and figure-ground organization are reversed, and, thus, two orangish peninsulas, going from the top to the bottom, connected to the mainland at the top, are now perceived (Fig. 4). Briefly, what appears as illusory colored and segregated in one figure is perceived as empty space without a clear coloration in the other figure and vice versa. Therefore, the peninsulas of Figs. 3 and 4 pop up as totally different objects not referable to the same juxtaposed contours.

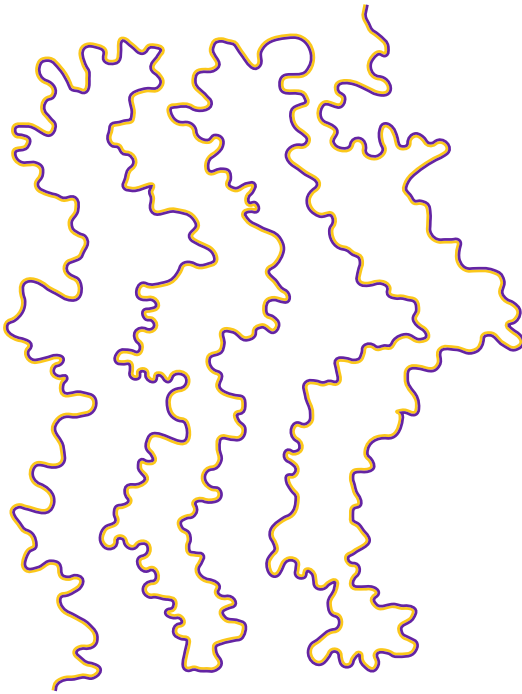


Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 3 Two undulated peninsulas emerging from the bottom are perceived illusorily colored by a light veil of orange tint spreading from the orange edges

Coloration Effects in the Watercolor Illusion

The phenomenology of the coloration effect in the watercolor illusion reveals the following main attributes:

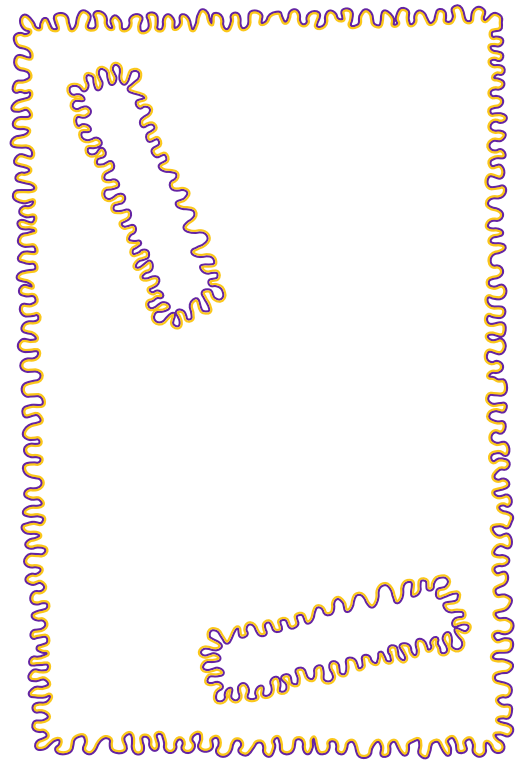
- (i) As in neon color spreading, the illusory color is approximately uniform. As shown in Fig. 5, the coloration, within the regions where the orange contours are closer, is the same as within the regions where they are distant.
- (ii) The coloration extends up to about 45.
- (iii) It is complete by 100 ms.
- (iv) Similarly to neon color spreading, all chromatic colors can generate the illusory coloration, as shown in Fig. 6, where an undefined irregular peninsula appears filled with a light blue tint. It should be noted that



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 4 By reversing the purple and orange contours of Fig. 3, the coloration and the figure-ground segregation is reversed: the two peninsulas going from the top to the bottom and connected to the mainland at the top

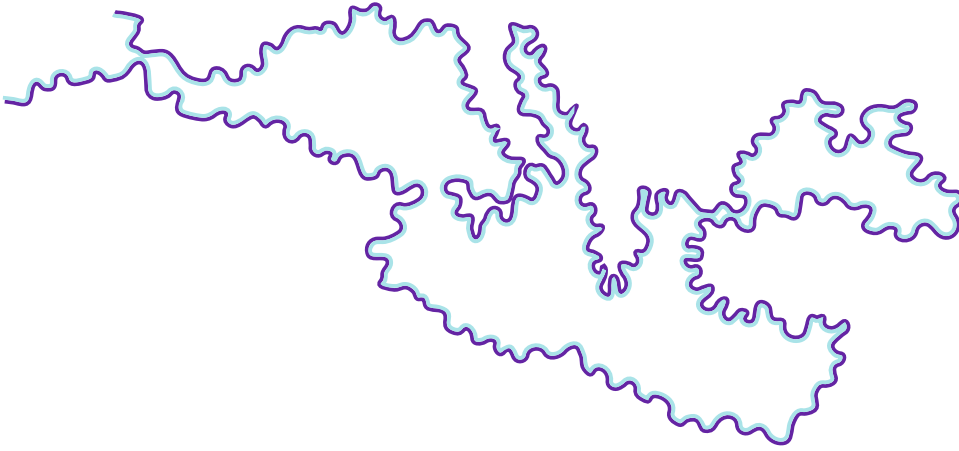
this peninsula is the Mediterranean Sea when the two adjacent chromatic contours are reversed. The watercolor illusion can also be induced with achromatic contours.

- (v) The coloration occurs on colored and black backgrounds. In Fig. 7, an undefined irregular peninsula (the Mediterranean Sea when the contours are reversed) appears filled with a purple tint.
- (vi) The optimal contour thickness is approx. 6 arcmin.
- (vii) The effect is stronger with wiggly contours, but it also occurs with straight contours and with chains of dots as shown in Fig. 8 [5, 6].
- (viii) High luminance contrast between inducing contours shows the strongest coloration effect; however, the color spreading is clearly visible at near equiluminance (see Fig. 9) [5, 6].

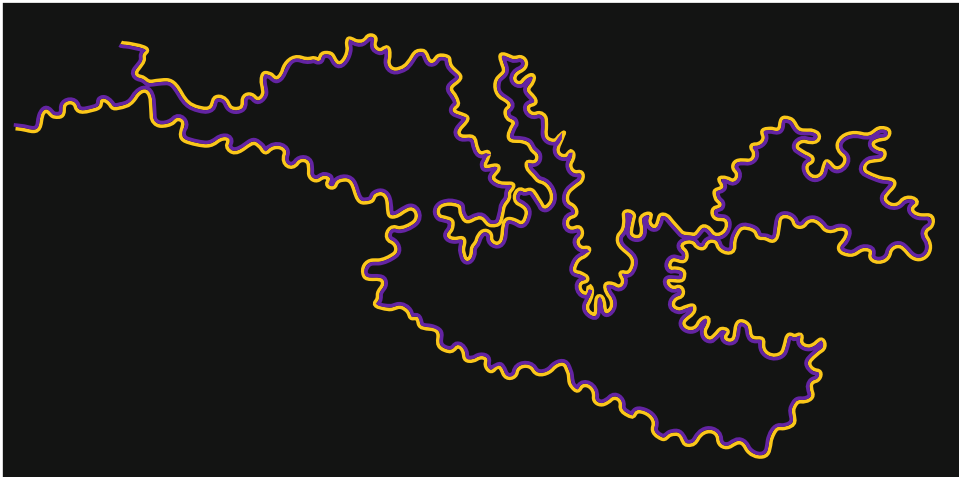


Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 5 The apparent coloration is approximately uniform, i.e., the coloration, within the regions where the orange contours are closer, is the same as the regions where they are distant

- (ix) The contour with a lower luminance contrast relative to the background spreads proportionally more than the contour with a higher luminance contrast.
- (x) The color spreads in directions other than the contour orientation.
- (xi) By reversing the colors of the two adjacent contours, the coloration reverses accordingly.
- (xii) Phenomenally, the coloration appears solid, impenetrable, and epiphanous as a surface color.
- (xiii) Similarly to neon color spreading [1–3], the watercolor illusion induces a complementary color when one of the two juxtaposed contours is achromatic and the other chromatic (see Fig. 10) [5]. The inside of the zigzagged annulus appears yellowish.



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 6 All the colors can generate the coloration effect of the watercolor illusion: the undefined irregular peninsula appears filled with a light blue tint



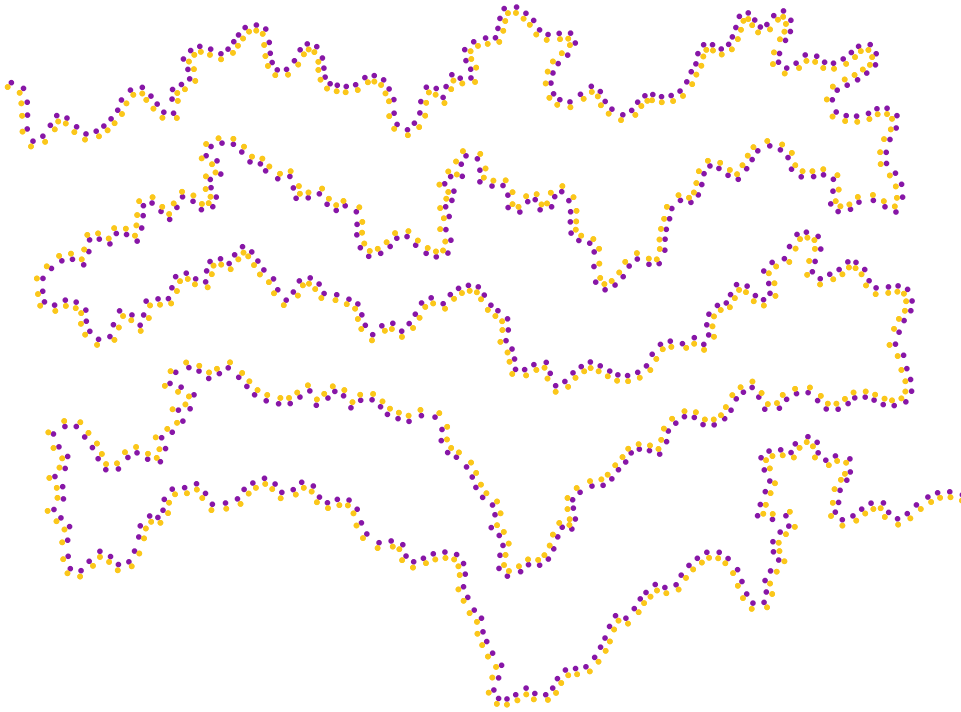
Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 7 The coloration occurs on colored and black backgrounds: an undefined irregular peninsula appears filled with a purple tint

Figural Effects in the Watercolor Illusion

The main properties of the figural effect are the following:

- (i) The watercolor illusion strongly enhances the “unilateral belongingness of the boundaries” [6, 7, 10].
- (ii) As in neon color spreading, the figural effect of the watercolor illusion is clearly perceived although it occurs in a different mode of appearance. The figure shows strong depth

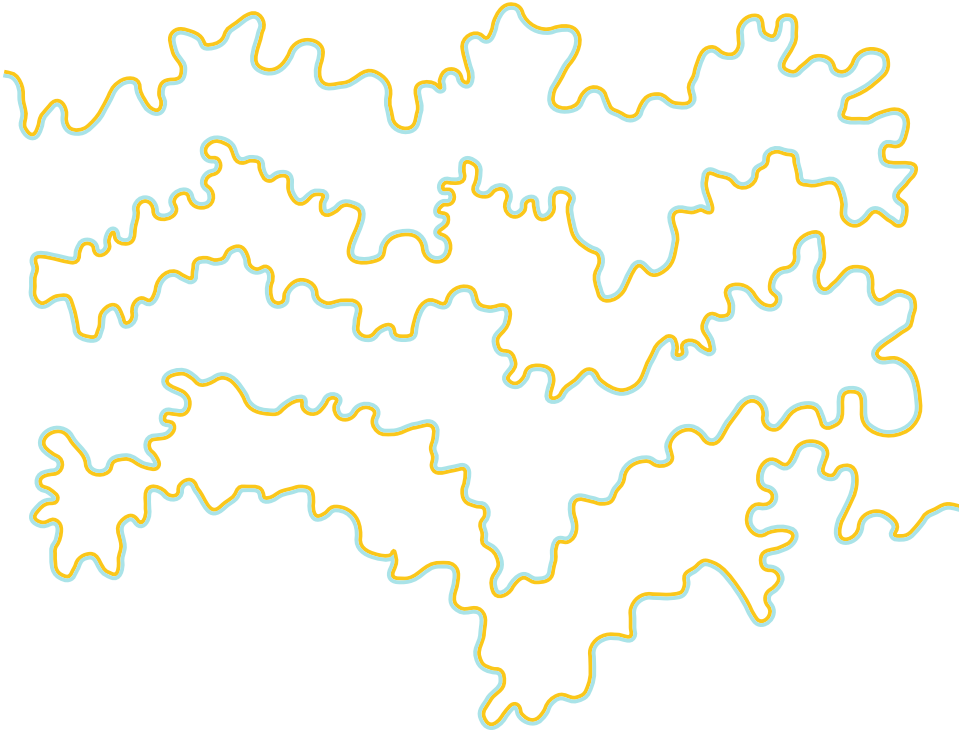
segregation and a volumetric rounded and three-dimensional attribute, while the perceived variation of color, going from the boundaries to the center of the object, is seen as a gradient of shading, as if light were reflected onto a volumetric and rounded object. Figure 11 shows undefined, rounded, and volumetric shapes differing from one row to another on a shapeless empty space due to the unilateral belongingness of the boundaries. The stars are totally invisible.



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 8 The watercolor illusion is stronger with wiggly contours, but it also occurs with chains of dots

- (iii) By reversing the colors of the two adjacent contours, the figure-ground segregation reverses accordingly. In Fig. 12, the same elements of Fig. 11, illustrated with reversed purple-orange contours, appear like juxtaposed stars. The undefined shapes differing from one row to another are now invisible.
- (iv) Under the previous conditions, the figure-ground segregation is not reversible and unequivocal.
- (v) The watercolor illusion determines grouping and figure-ground segregation more strongly than the Gestalt principles of proximity, good continuation, *Prägnanz*, relative orientation, closure, symmetry, convexity, past experience, similarity, surroundedness, and parallelism [6, 7, 10]. In Fig. 13, some examples showing the watercolor illusion respectively against and in favor of surroundedness, relative orientation, good continuation, past experience, and parallelism are illustrated.
- (vi) By reversing the luminance contrast of the background, e.g., from white to black, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses (Fig. 14) [10]. Going from the bottom to the top of the figure, the crosses become stars. These results are in contrast to Gestalt claim that the currently figural region is maintained even on black/white reversal.

This suggests that the watercolor illusion includes a new principle of figure-ground segregation, the asymmetric luminance contrast principle, stating that, all else being equal, given an asymmetric luminance contrast on both sides of a boundary, the region whose luminance gradient is less abrupt is perceived as a figure relative to the complementary more abrupt region, which is perceived as a background [10].



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 9 High luminance contrast between inducing contours shows the strongest coloration effect;

however, the color spreading is clearly visible at near equiluminance

Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 10 The inside of the zigzagged annulus appears yellowish



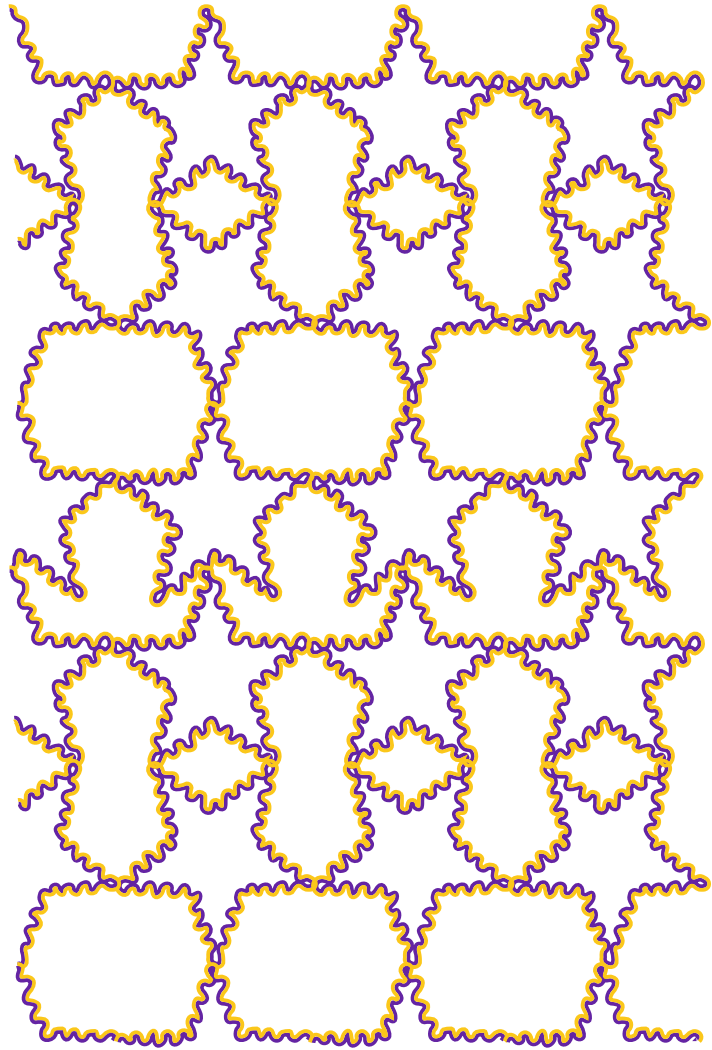
Similarities and Differences in Between the Two Illusions

By summing up the phenomenology of coloration and figural effects in both neon color spreading

and watercolor illusion, the former differs from the latter in both the appearance of the coloration (respectively, transparent vs. solid and impenetrable and diaphanous vs. epiphanous) and in the figural effects (respectively, transparent vs.

Color Spreading, Neon Color Spreading, and Watercolor Illusion,

Fig. 11 Watercolored undefined shapes differing from one row to another



opaque and dense and appearance as a “light,” a “veil,” a “shadow,” or a “fog” vs. rounded, thick, and opaque surface bulging from the background).

In spite of these differences, the two illusions are phenomenally similar in their clear color spreading and depth segregation. It is suggested [5] that while the similarities may depend on the local nearby transition of colors, equivalent in both illusions, the differences may be attributed to the global geometrical boundary conditions, dissimilar in the two illusions. As a matter of fact, while the neon color spreading is elicited by the continuation in the same direction of two contours of different

colors, the watercolor illusion occurs through their juxtaposition.

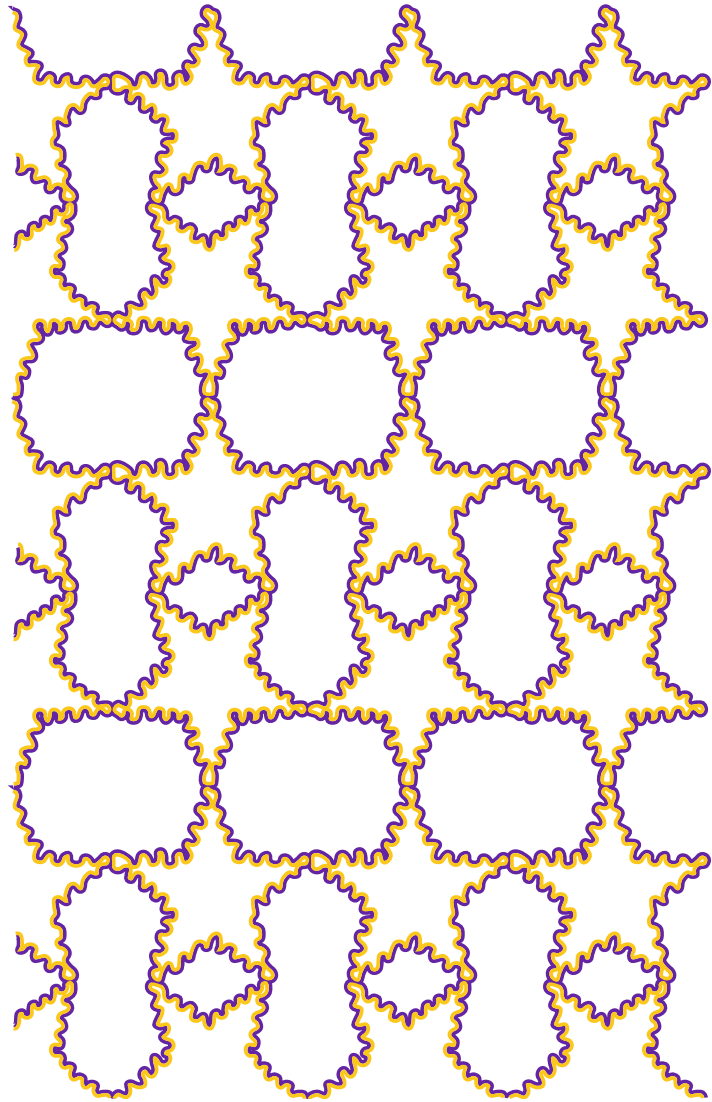
If this is true, the phenomenal differences between the two illusions can be reduced or eliminated through geometrical variations that bring both phenomena to a common limiting case placed in between and based on the local nearby transitions of colors.

A Limiting Case

Figure 15 shows four conditions that gradually introduce the limiting case. Figure 15a illustrates a neon color spreading that represents the starting

Color Spreading, Neon Color Spreading, and Watercolor Illusion,

Fig. 12 The same elements of Fig. 11, illustrated with reversed purple-orange contours, appear like juxtaposed stars



condition where concentric purple arcs continue by becoming orange. Now, the resulting inset square annulus appears as a transparent veil of orange color not like a ghostly circular veil of translucent color as in Fig. 1. This difference in the color and figural appearance is likely related to the high contrast between the two inducing colors.

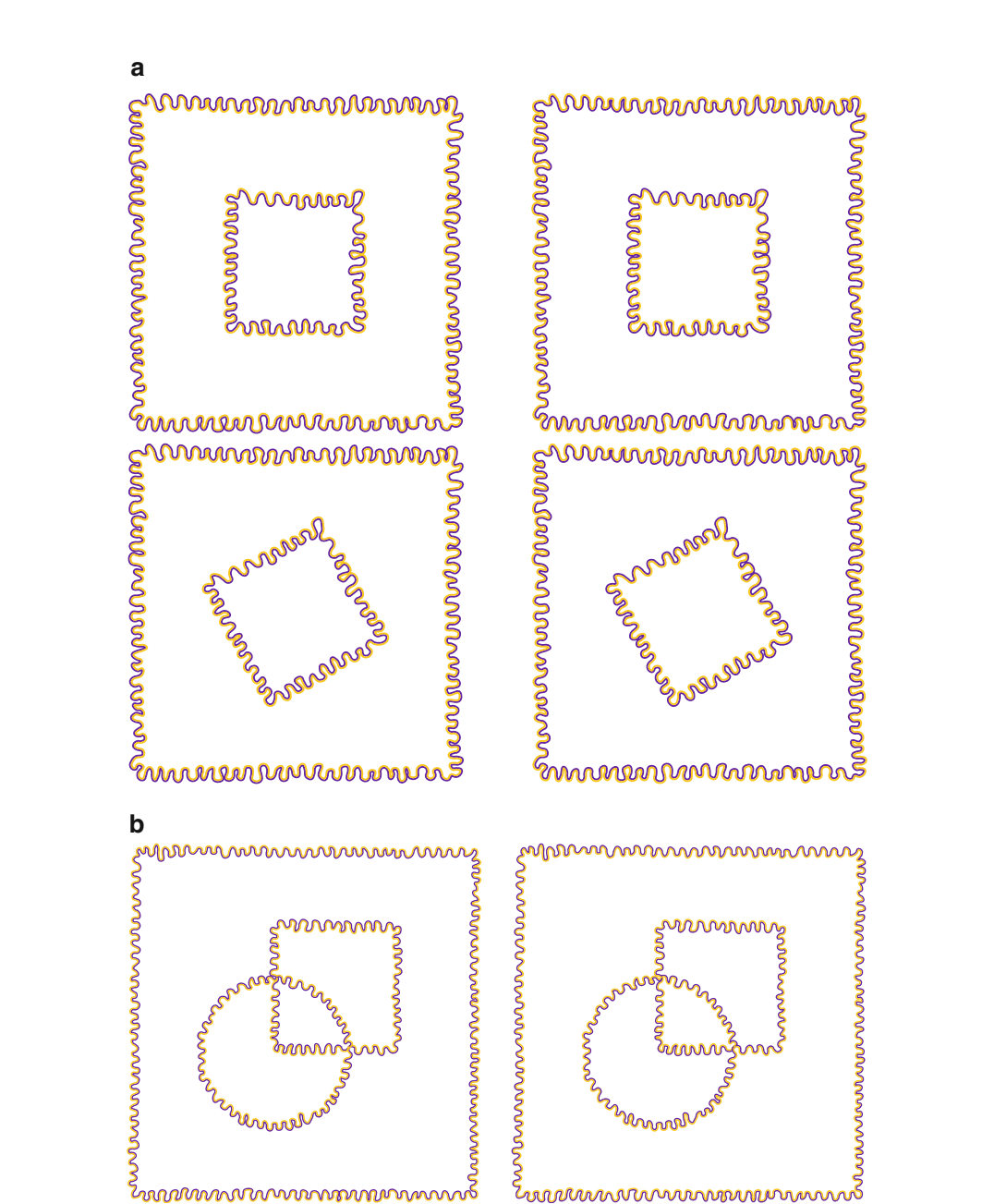
Gradual steps toward the final combination of the two illusions in a limiting case are illustrated in Fig. 15b, c. Geometrically, in Fig. 15b, the orange inset arcs are reduced to short dashes, creating a condition in between neon color spreading and the watercolor illusion: from the neon

color spreading perspective, the inducing elements are contours *continuing* in short dashes (or elongated dots), but from the watercolor perspective, the terminations of the inducing arcs contain *juxtaposed* short dashes. Under these conditions, a coloration effect, not weaker than that of Fig. 15a, is perceived. However, it manifests a poor diaphanous and surface appearance. The illusory figure appears as a fuzzy square annulus, yellowish and brighter than the back-ground. It is worthwhile to note that the further reduction of the dashes to dots does not change significantly the strength of these effects.

The geometrical reduction in between neon color spreading and the watercolor illusion and opposite to the one of Fig. 15b is illustrated in Fig. 15c. Under these further conditions, all else being equal, short dashes become the purple arcs

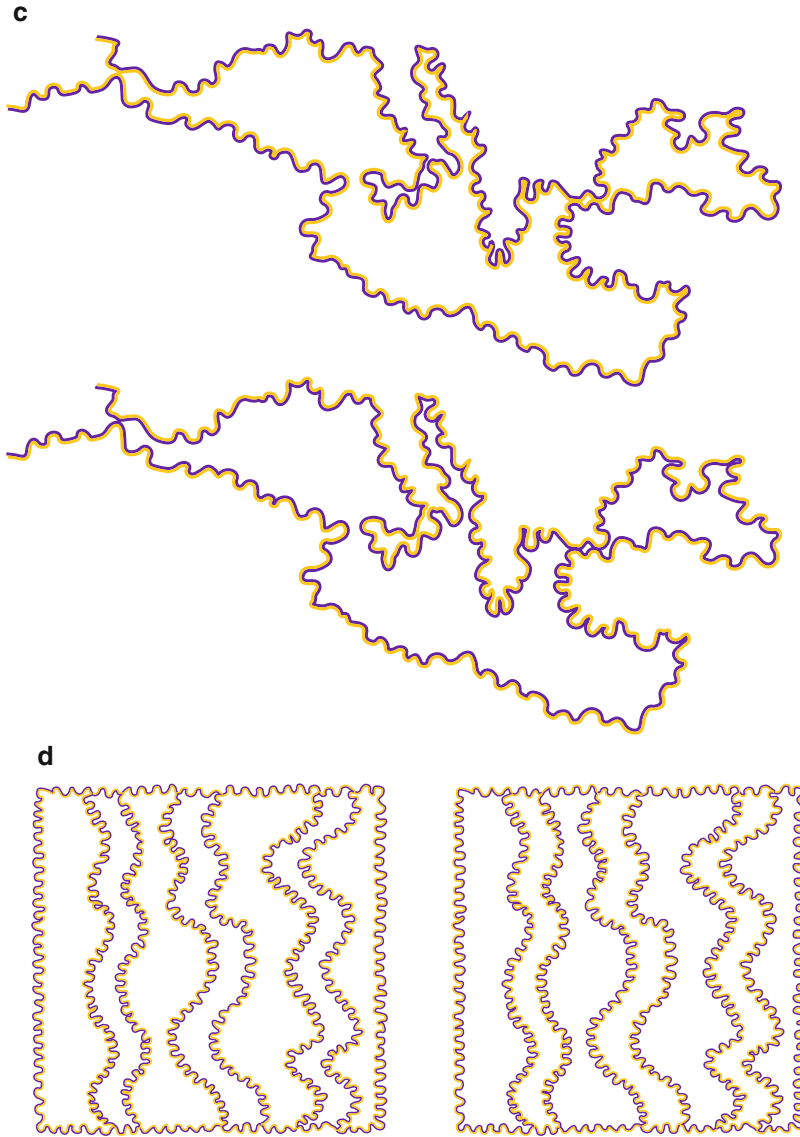
of Fig. 15a. Now the coloration effect is weaker than that of Fig. 15a.

Given these geometrical prerequisites, the final step toward the limiting case becomes immediate and consists in putting together the previous



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13 Some examples showing the watercolor illusion respectively against and in favor of

surroundedness, relative orientation, good continuation, past experience, and parallelism



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13 (continued)

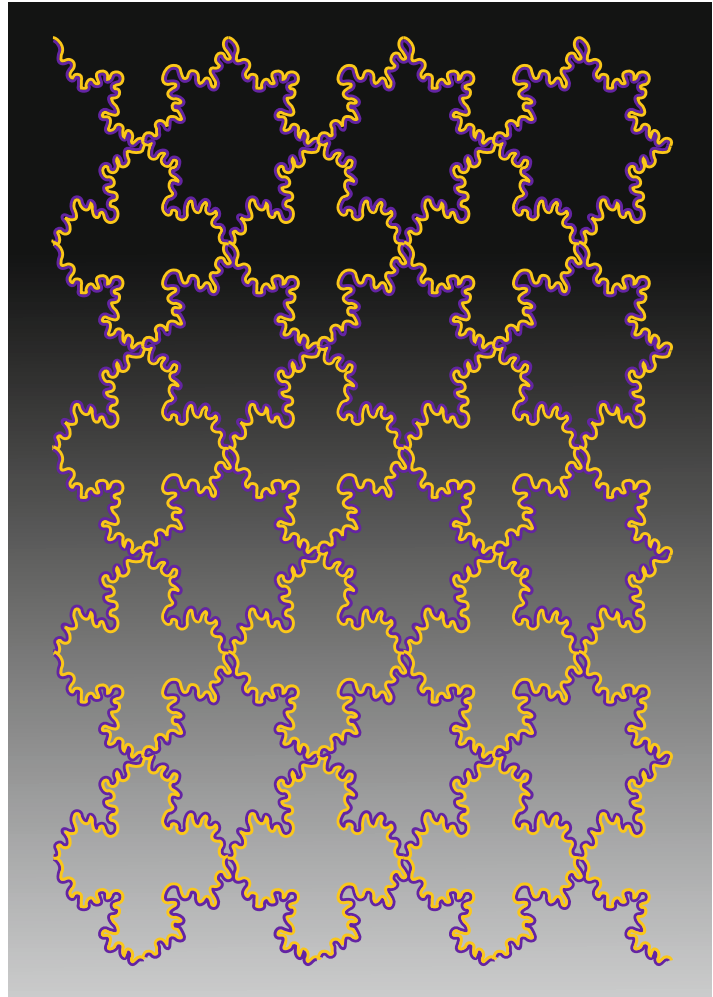
opposite reductions as shown in Fig. 15d. The results show that by reducing both the purple and orange arcs of Fig. 15a to short dashes, the coloration and figural effects do not change significantly [5]. This is corroborated by previous outcomes according to which the watercolor illusion occurs not only by using juxtaposed lines but also by using juxtaposed chains of dots [6, 7, 10]. Under these conditions both coloration and figural effects

become weaker and weaker as the density of the dots becomes sparser and sparser.

The two-dot juxtaposition of Fig. 15d can be considered as a true limiting case for neon color spreading and the watercolor illusion. As a matter of fact, (i) the two-dot limiting case can be considered as the geometrical common condition, beneath. (ii) The strength of both coloration and figural effects does not change

Color Spreading, Neon Color Spreading, and Watercolor Illusion,

Fig. 14 By reversing the luminance contrast of the background, from white to black, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses: going from the bottom to the top of the figure, the crosses become stars

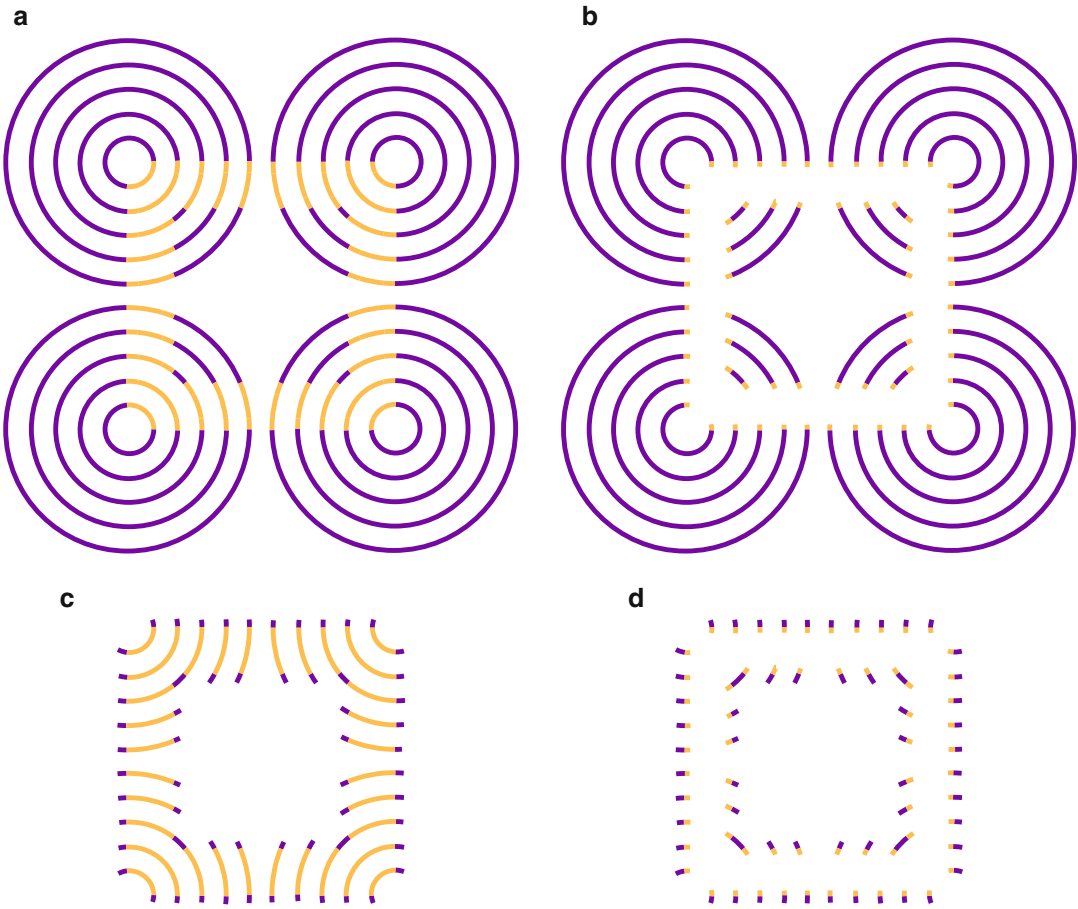


significantly; therefore, the specific mode of appearance of coloration and figural effects in the two illusions is elicited by different local and global distributions of nearby transitions of colors that, in their turn, induce different boundary organizations. (iii) Phenomenally, the differences between the two illusions, where the inner changes are based on continuation and juxtaposition of contours, can now be reconsidered and unified in terms of transition. This is not only a linguistic alternative but also it can bring advantages by providing support for a simple common neural model. (iv) The limiting case can suggest variations of the two illusions that manifest coloration and figural attributes in between the neon color spreading

and the watercolor illusion, as shown in the next section.

Near the Limiting Case

By increasing the width of one of the two juxtaposed contours of the watercolor illusion to such an extent that the contour becomes a surface, the watercolor illusion manifests geometrical properties similar to the neon color spreading and, as a consequence, also shows different coloration and figural effects. The resulting coloration does not assume surface color properties, but properties more similar to the neon color spreading. It appears diaphanous like a foggy coloration



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 15 (a) The neon color spreading defined by the continuation of lines of different color; (b) a condition in between neon color spreading and

watercolor illusion, where the orange inset arcs are reduced to short dashes; (c) a condition in between the two illusions, where the purple surrounding arcs of (a) are reduced to short dashes; (d) the two-dots limiting case

diffusing everywhere in the background or as a colored light (Fig. 16).

Similarly to the previous condition, the coloration effect of Fig. 17 gives to the illusory star a fuzzy luminous quality. While in Fig. 16 the coloration belongs to the background, in Fig. 17 it belongs to the figure; however, the star does not manifest a strong surface appearance, but its inner surface appears brighter and yellowish and foggy and smooth.

Taken together, these figures suggest that (i) the modes of appearance of coloration are strongly related to boundary conditions that induce specific figural effects; (ii) by changing the boundary conditions, coloration and figural

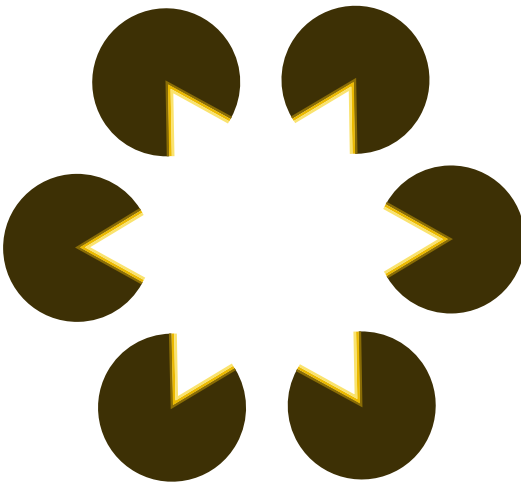
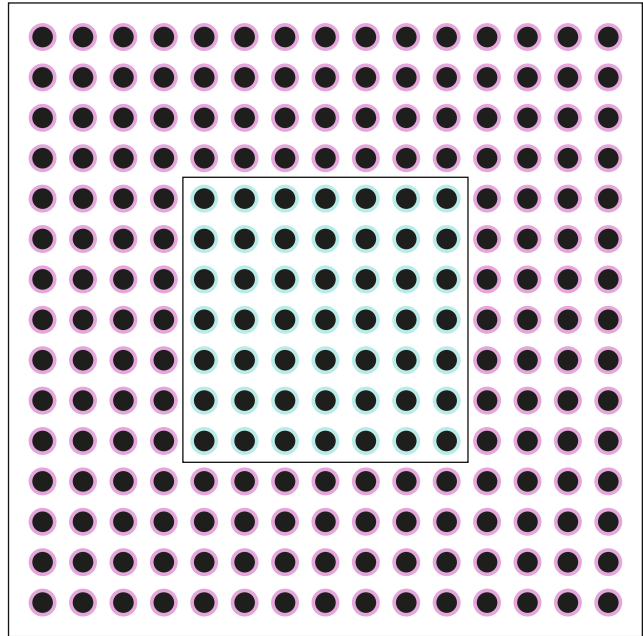
attributes are perceived more similar to one, to the other illusion, or in between; and (iii) given this variety of appearances on the basis of different conditions, a simpler set of boundary cases, like in the limiting case, can unify both effects using local transitions of colors and can help to explain similarities and dissimilarities of the two illusions.

Neural Mechanisms Underlying the Two Illusions

On the basis of the previous results, coloration and figural effects may derive from parallel processes.

Color Spreading, Neon Color Spreading, and Watercolor Illusion,

Fig. 16 A light blue coloration filling in the inset square appears surrounded by a red spreading. The coloration effect is not accompanied by a figural effect with a plain volumetric property, but it appears diaphanous like a foggy veil of color



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 17 The illusory coloration of the star appears fuzzy and luminous, and it manifests a poor surface appearance

At a feature processing stage, the short-range interaction area around and in between the two dots produces the color spreading common to both illusions, and at a parallel boundary processing stage, the different geometrical structures in both illusions organize the color spreading to elicit different figural effects. Moreover, the

reduction of the neon color spreading and watercolor illusion to a common limiting case can suggest a common and an easier explanation that can be based on the FACADE neural model of biological vision [5]. The model posits that two processes, boundary grouping and surface filling-in [11, 12] substantiated by the cortical interblob and blob streams, respectively, within cortical areas V1 through V4, are responsible of how local properties of color transitions activate spatial competition among nearby perceptual boundaries, with boundaries of lower-contrast edges weakened by competition more than boundaries of higher-contrast edges. This asymmetry induces spreading of more color across these boundaries than conversely. These boundary and surface processes show complementary properties that can also predict how depth and figure-ground effects are generated in these illusions.

Other related findings to both illusions [13–15] showed that neurons in V2 respond with different strength to the same contrast border, depending on the side of the figure to which the border belongs, implying a neural correlate process related to the unilateral belongingness of the boundaries. Figure-ground segregation may be processed in areas V1 and V2, in inferotemporal cortex and the

human lateral occipital complex. Also the color spreading of the two illusions might have its explanation in the cortical representation of borders [9].

Summary

The color spreading is a long-range assimilative spread of color emanating from a thin-colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. Two main examples of color spreading are the well-known neon color spreading and the watercolor illusion. The coloration and the figural properties of the two illusions, studied using phenomenal and psychophysical observations, can be reduced to a common limiting condition, i.e., a nearby color transition called the “two-dots limiting case,” which explains their perceptual similarities and dissimilarities and suggests a common explanation.

Cross-References

- [Assimilation](#)
- [Color Phenomenology](#)
- [Complementary Colors](#)
- [Katz, David](#)
- [Perceptual Grouping and Color](#)

References

1. Varin, D.: Fenomeni di contrasto e diffusione cromatica nell'organizzazione spaziale del campo percettivo. *Riv. Psicol.* **65**, 101–128 (1971)
2. van Tuijl, H.F.J.M.: A new visual illusion: neon-like color spreading and complementary color induction between subjective contours. *Acta Psychol.* **39**, 441–445 (1975)
3. Bressan, P., Mingolla, E., Spillmann, L., Watanabe, T.: Neon colour spreading: a review. *Perception* **26**, 1353–1366 (1997)
4. Pinna, B., Brelstaff, G., Spillmann, L.: Surface color from boundaries: a new ‘watercolor’ illusion. *Vis. Res.* **41**, 2669–2676 (2001)
5. Pinna, B., Spillmann, L., Werner, J.S.: Anomalous induction of brightness and surface qualities: a new illusion due to radial lines and chromatic rings. *Perception* **32**, 1289–1305 (2003)
6. Pinna, B., Grossberg, S.: The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **22**, 2207–2221 (2005)
7. Werner, J.S., Pinna, B., Spillmann, L.: The brain and the world of illusory colors. *Sci. Am.* **3**, 90–95 (2007)
8. Pinna, B.: Watercolor illusion. *Scholarpedia* **3**, 5352 (2008)
9. von der Heydt, R., Pierson, R.: Dissociation of color and figure-ground effects in the watercolor illusion. *Spat. Vis.* **19**, 323–340 (2006)
10. Devinck, F., Knoblauch, K.A.: Common signal detection model for the perception and discrimination of the watercolor effect. *J. Vis.* **12**, 1–14 (2012)
11. Komatsu, H.: The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* **7**, 220–231 (2006)
12. Murakami, I.: Perceptual filling-in. In: *Encyclopedia of Neuroscience*. Springer, Berlin (2008)
13. Zhou, H., Friedman, H.S., von der Heydt, R.: Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611 (2000)
14. von der Heydt, R., Zhou, H., Friedman, H.S.: Neural coding of border ownership: implications for the theory of figure-ground perception. In: Behrmann, M., Kimchi, R., Olson, C.R. (eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, pp. 281–304. Lawrence Erlbaum Associates, Mahwah (2003)
15. Friedman, H.S., Zhou, H., von der Heydt, R.: The coding of uniform color figures in monkey visual cortex. *J. Physiol. Lond.* **54**, 593–613 (2003)

Color Statistics

- [Color Scene Statistics](#), [Chromatic Scene Statistics](#)

Color Stereoscopic Effect

- [Chromostereopsis](#)

Color Stereoscopia

- [Chromostereopsis](#)

Color Symbolism

► Chinese Color Language

Color Synaesthesia

► Color Synesthesia

Color Synesthesia

Berit Brogaard^{1,2}, Dimitria Electra Gatzia³ and Jennifer Matey⁴

¹Department of Philosophy, University of Oslo, Oslo, Norway

²Brogaard Lab for Multisensory Research, University of Miami, Miami, FL, USA

³Department of Philosophy, University of Akron, Akron, OH, USA

⁴Department of Philosophy, Southern Methodist University, Dallas, TX, USA

Synonyms

Color synaesthesia

Definition

Color synesthesia is a condition in which sensory or cognitive inducers elicit involuntary, atypical, concurrent color experiences.

Marks of Color Synesthesia

Synesthesia is a condition that involves unusual pairings across modalities. In color synesthesia, a sensory or cognitive stimulus such as a grapheme or a sound automatically and involuntarily induces specific color experiences. Various types of color synesthesia have been identified,

including calendar-color synesthesia, sound-color synesthesia, taste-color synesthesia, and even fear-color synesthesia. However, the most prevalent form of the condition is grapheme-color synesthesia, where numbers or letters induce highly specific color experiences [1].

One of the marks of grapheme-color synesthesia is that it exhibits test-retest reliability, meaning that the synesthetic colors subjects identify relative to specific stimuli in the initial testing phase are nearly identical to those identified in retesting phases [1] (see Fig. 1).

Another mark of grapheme-color synesthesia is its uniqueness: synesthetic inducer-concurrent pairs are unique for each individual [1, 2]. For example, the letter A may induce a synesthetic experience of redness in one grapheme-color synesthete but a synesthetic experience of blueness in another. In fact, each grapheme has been found to trigger each of the 11 Berlin and Kay colors in different grapheme-color synesthetes (i.e., red, pink, orange, yellow, green, blue, purple, brown, black, white, gray). Another aspect of the uniqueness of the grapheme-color synesthesia is how the synesthetic colors are experienced [3]. Projector grapheme-color synesthetes see synesthetic colors as projected out onto the world. Pop-up effects are taken to be signifiers of projector synesthesia. Associator grapheme-color synesthetes, by contrast, experience colors in the mind's eye. Their synesthetic colors are triggered by automatic retrieval of mnemonic associations, which are more akin to mental imagery than to remembering something learned in childhood [4]. Both projector and associator grapheme-color synesthesias involve the triggering of a highly specific and automatic inducer such as a colored letter or number.

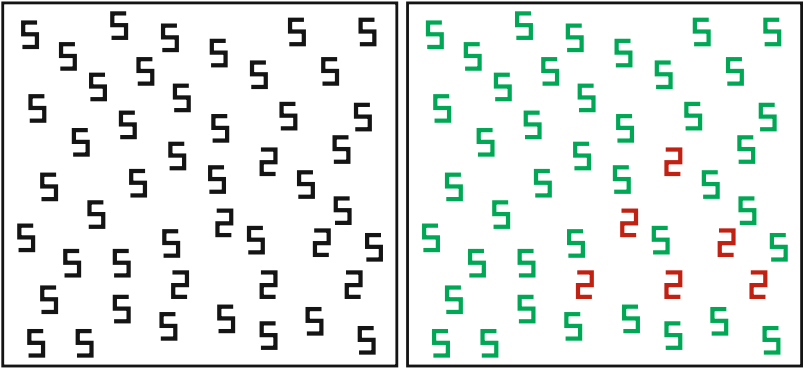
Notwithstanding the individual differences among grapheme-color synesthetes, studies indicate that synesthetes show significant group preferences for certain grapheme-color associations [5]. For example, grapheme-color synesthetes tend to associate the letter *a* with red, the letters *i* and *o* with black or white, and the letter *n* with brown. These associations are not entirely random. While studies suggest that the experiences between grapheme-color synesthetes and non-

Age/graph	0	1	2	3	4	5	6	7	8	9
3	/	B	Y	G	P	R	Bl	W	Br	R
4	/	B	Y	G	P	R	Bl	W	Br	R
5	Go	B	Y	G	P	R	DBr	W	Br	R
6	Go	B	Y	G	P	R	DBr	W	Br	R
7	B	B	Y	G	P	R	Br	W	Br	R
8	B	B	Y	G	P	R	Bl	W	Br	R

Color Synesthesia, Fig. 1 Example of test-retest reliability of synesthetic experience in one of the St. Louis Synesthesia Lab’s associator grapheme-color synesthetes

from ages 3 to 8 (Go = gold, B = blue, Y = yellow, G = green, P = purple, R = red, Bl = black, DBr = dark brown, Br = brown, W = white)

Color Synesthesia, Fig. 2 When normal subjects are presented with the figure on the *left*, it takes them several seconds to identify the hidden shape. Some grapheme-color synesthetes instantly see the *triangular* shape because they experience the 2s and the 5s as having different colors



synesthetes differ significantly, there are some common grapheme-color preferences among the two groups [6]. For example, both grapheme-color synesthetes and (forced- or free-choice) controls associate the letter *f* with green, the letter *b* with blue, and the letter *v* with purple.

High-Level Versus Low-Level Synesthesias

An open question about grapheme-color synesthesia is whether it is triggered by low-level perceptual features (such as the curvature of the number 3) or high-level perceptual features of inducers (such as concepts inherent in a grapheme). Collectively, the evidence seems to indicate that both low- and high-level synesthesias occur in different individuals [4].

Some grapheme-color synesthetes seem to experience a pop-out effect in visual search experiments. For example, if a cluster of 2s is embedded

in an array of randomly placed 5s, normal subjects take several seconds to find the shape formed by the 2s, whereas for some grapheme-color synesthetes, 5s pop-out instantly (see Fig. 2) [7].

Pop-out effects seem to indicate that some grapheme-color synesthesias are triggered by low-level perceptual features because they involve segregation. Specifically, grapheme-color synesthetes are able to experience the targets and the distractors as distinct stimuli: the 5s (which function as the target) are seen as forming a triangle and appear to be distinct from the 2s (which function as the distractors). Since only perceptual features processed early in visual processing can lead to segregation, pop-up effects seem to provide support for low-level grapheme-color synesthesias [7]. Pop-out effects are also taken to be evidence for projector grapheme-color synesthesias. Projector grapheme-color synesthetes see synesthetic colors as projected out onto the world. The induced colors (in Fig. 2) are experienced as belonging to the 2s and 5s

seen on the template; they are not experienced as mere associations.

Visual search experiments that give rise to pop-out effects raise the question of whether synesthesia requires focused attention. If an inducer elicits synesthetic colors without it being attended, then a target (e.g., a 5 in Fig. 2) that has a distinct synesthetic color from a distractor (e.g., a 2 in Fig. 2) should capture attention and lead to highly efficient identification. But if an inducer does not elicit synesthetic colors until it is attended, then the identification would be relatively inefficient.

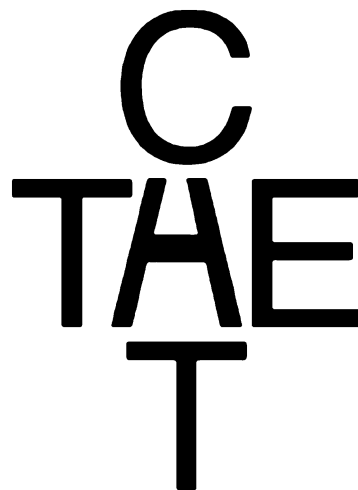
Although some findings appear to suggest that pop-out effects occur preattentively, upon closer investigation, the current evidence indicates that an inducer does not elicit synesthetic colors until it is attended. One study [8] used a variation on the standard visual search experiment to test a subject's (J) search efficiency. J was shown an array of black graphemes on a colored background, some of which induced synesthetic colors. The colored background was either congruent or incongruent with the synesthetic color of the target. The researchers found that J was more efficient in her search when the background and the synesthetic color were incongruent than when they were congruent. This indicates that synesthetic colors attracted attention only when they were clearly distinct from the background. In another study, a subject, PM, was able to identify graphemes quickly but only when the target was within a few degrees of visual angle from fixation, indicating that PM's synesthesia is within the focus of attention [9].

Other findings seem to nevertheless indicate the presence of low-level synesthesia. For example, one study found that the strength of the induced colors in one grapheme-color synesthete (JC) varied depending on whether the graphemes were presented in high or low contrast [10]. JC's synesthetic colors were reduced or absent at low contrasts but not at high contrasts. Since JC's synesthesia is elicited by low-level features (i.e., high contrast), this study indicates the presence of low-level synesthesia.

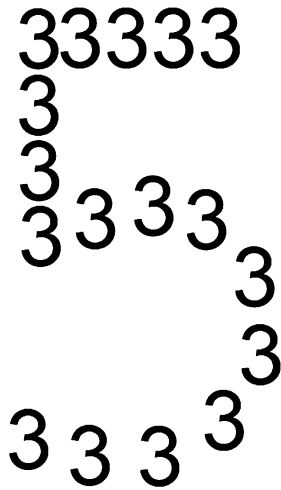
All of the aforementioned studies indicate the presence of low-level synesthesia. However, studies also seem to provide support for high-level

synesthesias. In one experiment, a single grapheme induced different synesthetic colors depending on the context, e.g., the category or concept associated with the inducer. In one subject, the middle letter "H" induced a red synesthetic experience when she read the word shown in Fig. 3 as "CAT" but a brown synesthetic experience when she reads the word as "THE" [7].

In a similar experiment, color-grapheme synesthetes (JC and ER) were able to switch between their synesthetic colors when they were shown Navon-type figures such as an image of a large 5 constructed using smaller 3s (Fig. 4) [11]. For example, attending to the entire figure (i.e., the number 5) induced one synesthetic color, but attending to its parts (i.e., the smaller 3s comprising the 5) induced another. Similarly, when a display that could be viewed either as the letters I and V placed next to each other or as the Roman numeral four, synesthetic colors were induced only in the former case [7]. The fact that the very same grapheme can induce different synesthetic colors depending on the context in which it occurs suggests that some grapheme-color synesthesias are triggered by high-level perceptual features. It is possible that such contextual effects associated with high-level synesthesia are the result of top-



Color Synesthesia, Fig. 3 Synesthetes interpret the middle letter as an A when it occurs in "CAT" and as an H when it occurs in "THE." The color of their synesthetic experience will depend on whether the grapheme is considered part of the word "CAT" or "THE"



Color Synesthesia, Fig. 4 Some synesthetes are able to switch between their synesthetic colors when they were shown Navon-type figures such as an image of a large 5 constructed using smaller 3s. Two different synesthetic colors are induced depending on whether synesthetes attend to the entire figure, the number 5, or it's parts, the 3s that comprise the 5

down modulation on early visual processing [7]. If so, these cases may turn out to be cases of low-level synesthesia.

Cognitive and Neural Mechanisms

The precise cognitive and neural mechanisms underlying color synesthesia are unknown. Some propose cognitive models of synesthesia based on the modularity thesis about processing. The modular theory holds that functionally individuated and introspectively opaque, encapsulated subunits of the mind process information in a way akin to a computer. One hypothesis is that synesthesia occurs when there is a breakdown in encapsulation between modules that process color and whatever other sensory or semantic features are involved [5, 12]. Others contend that synesthesia might result from the presence of an extra-module whose function is the mapping from a sensory or semantic input to a cross modal output [13].

A second type of account appeals more directly to the neural mechanisms underlying synesthesia. Structural connectivity mechanisms have been

proposed for standard cases of grapheme-color synesthesia and sound-color synesthesia, suggesting unusual connectivity between color areas in the visual cortex and the adjacent visual word form area or the auditory cortex [7, 10]. Evidence for this hypothesis comes from brain-imaging studies showing enhanced anatomical connectivity near the implicated brain regions in grapheme-color synesthesia [14] and sound-color synesthesia [15].

Functional connectivity mechanisms have also been proposed for color synesthesia, according to which neural networks that give rise to synesthetic experiences exhibit excessive disinhibition or hyperexcitement of neurotypical connections through a change in neurotransmitters [16]. On one version of this hypothesis, color synesthesia is due to disinhibited feedback from an area of the brain that binds information from different sensory modalities [17]. Evidence for this hypothesis comes from an analogous case in which a patient PH reported seeing visual movement in response to tactile stimuli following acquired blindness [18]. As PH was blind, he could not have received the information via standard visual pathways. It is plausible that PM's synesthetic experience results from disinhibited feedback from brain regions that receive information from other senses.

The fact that synesthetic experiences can arise when subjects are under the influence of psychedelics provides further evidence for functional connectivity mechanisms [19, 20]. It has been shown in several studies that psychedelic hallucinogens that function primarily as serotonin agonists, such as psilocybin, LSD, and mescaline, often induce transient synesthesia, presumably through an alteration of functional brain connectivity [19, 21]. It is unknown, however, whether drug-induced synesthesia and developmental synesthesia have the same underlying mechanism.

On another version of the hypothesis that synesthesia is due to aberrant functional connectivity, color synesthesia arises as a result of aberrant reentrant processing [22]. This proposal is akin to the disinhibited feedback hypothesis but suggests specifically that high-level information reenters color areas in the visual cortex and that it is this form of reentrant information processing that

leads to the experience of synesthetic colors. This model would explain why visual context and meaning typically influence which synesthetic colors a grapheme gives rise to [2].

Yet another hypothesis is that some forms of color synesthesia are learned in early childhood, for example, as a result of repeated exposure to refrigerator magnets [4]. Grapheme-color binding, for example, may in some cases be the result of enhanced activation of color areas in the visual cortex in response to automatically triggered semantic memory recall [23–25]. Whether color synesthesia can be acquired in adulthood through similar perceptual learning mechanisms remains to be seen.

It is plausible that different forms of color synesthesia proceed via different mechanisms. Cases of color synesthesia have been reported in which the visual cortex is not involved in generating synesthetic colors [26]. None of the aforementioned hypotheses, despite their plausibility in run-of-the mill cases, can explain more unusual cases of color synesthesia.

Cognitive Advantages of Color Synesthesia

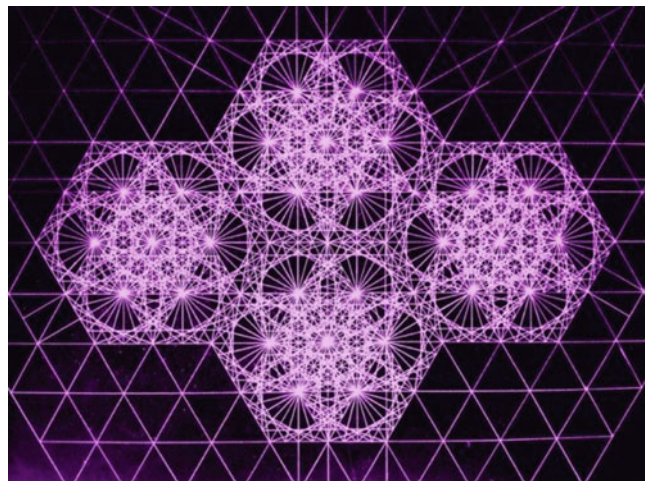
Various cognitive advantages have been claimed for synesthesia. Some case studies suggest that

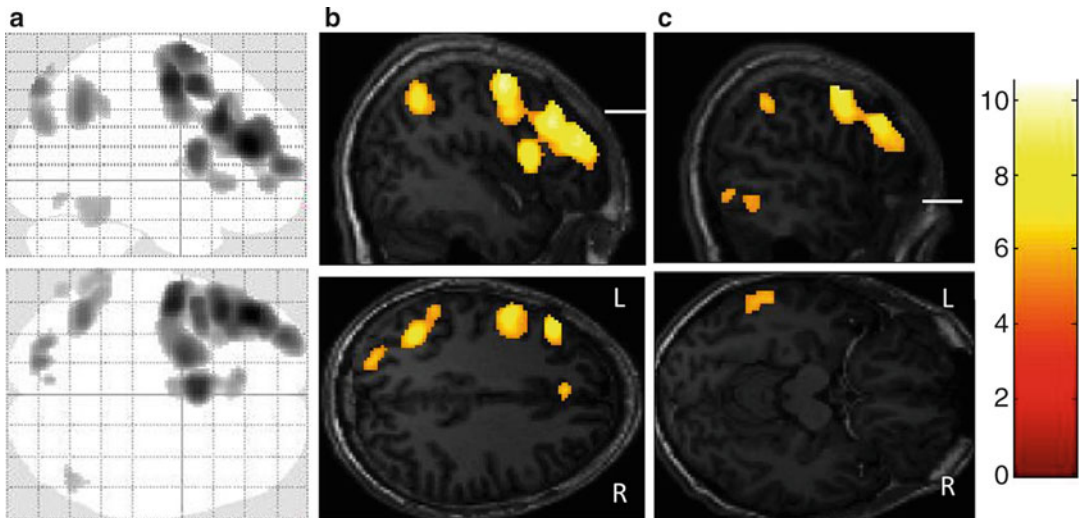
grapheme-color synesthetes may have greater recall ability for digits and written names when compared to non-synesthetes [27].

In rare cases, color synesthesia has been associated with extreme mathematical skills. Subject DT, for example, sees numbers as three-dimensional colored, textured forms [28]. His synesthesia gives him the ability to multiply high numbers very rapidly. He reports that the product of multiplying two numbers is the number that corresponds to the shape that fits between the shapes corresponding to the multiplied numbers. Subject DT's color synesthesia also gives rise to extreme mnemonic skills. DT currently holds the European record in reciting the decimal points of the number pi (π). An fMRI study comparing DT to controls while attempting to locate patterns in number sequences indicated that DT's synesthetic color experiences occur as a result of information processing in nonvisual brain regions, including temporal, parietal, and frontal areas [28]. This advantage may be facilitated by top-down modulation and has been used to support the view that photisms represent letterocity or numerocity [29].

Another subject, JP, was found to have exceptional abilities to draw complex geometrical images by hand and a form of acquired synesthesia for mathematical formulas and moving objects, which he perceives as colored, complex geometrical figures [26] (see Fig. 5).

Color Synesthesia,
Fig. 5 Image hand-drawn
by subject JP





Color Synesthesia, Fig. 6 Sagittal slices. Activation induced by the image-inducing formula contrasted to non-inducing formulas. The SPM(T) maps were

thresholded at family-wise-error-corrected p -value 0.01 and overlaid on JP's structural T1-weighted MRI which was standardized into MNI-space using SPM8 [26]

JP's synesthesia began in the wake of a brutal assault that led to unspecified brain injury. A fMRI study contrasting activity resulting from exposure to image-inducing formulas and non-inducing formulas indicated that JP's colored synesthetic images arise as a result of activation in areas in the temporal, parietal, and frontal cortices in the left hemisphere. The image-inducing formulas as contrasted with the non-inducing formulas induced no activation in the visual cortex or the right hemisphere [26] (see Figs. 5 and 6).

This case also suggests that at least some forms of color synesthesia can give rise to cognitive advantages in the area of mathematics. As the visual cortex does not appear to be directly involved in generating the synesthetic images, it also suggests that at least some forms of the condition are best characterized as forms of top-down modulation that proceeds via nonstandard mechanisms.

Enhanced creativity may be another advantage conferred by synesthesia. There is some evidence that synesthetes are on average more likely to work as artists [30, 31]. Both synesthesia and creativity involve linking unrelated ideas to reveal deep similarities, which may be due to the greater connectivity between processing areas [7].

Cross-References

- [Afterimage](#)
- [Ancient Color Terminology](#)
- [Appearance](#)
- [Chromatic Contrast Sensitivity](#)
- [Color and Visual Search, Color Singletons](#)
- [Color Constancy](#)
- [Color Contrast](#)
- [Color Phenomenology](#)
- [Color Processing, Cortical](#)
- [Color Psychology](#)

References

1. Cytowic, R.E., Eagleman, D.M.: *Wednesday Is Indigo Blue*. MIT Press, Cambridge, MA (2009)
2. Dixon, M.J., Smilek, D.: The importance of individual differences in grapheme-color synesthesia. *Neuron*. **45**, 821–823 (2005)
3. Dixon, M.J., Smilek, D., Merikle, P.M.: Not all synaesthetes are created equal: projector versus associator synaesthetes. *Cogn. Affect. Behav. Neurosci.* **4**, 335–343 (2004)
4. Witthoft, N., Winawer, J.: Learning, memory, and synesthesia. *Psychol. Sci.* **24**(3), 258–265 (2013)
5. Baron-Cohen, S., Harroson, J., Goldstein, L.H., Wyke, M.: Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception*. **22**, 419–426 (1993)

6. Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., et al.: Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cogn. Neuropsychol.* **22**(8), 1069 (2005)
7. Ramachandran, V.S., Hubbard, E.M.: The emergence of the human mind: some clues from synesthesia. In: Robertson, L.C., Sagiv, N. (eds.) *Synesthesia: Perspectives from Cognitive Neuroscience*, pp. 147–190. Oxford University Press, Oxford (2005)
8. Smilek, D., Dixon, M.J., Merikle, P.M.: Synaesthetic photisms guide attention. *Brain Cogn.* **53**, 364–367 (2003)
9. Laeng, B., Svardal, F., Oelmann, H.: Does color synesthesia pose a paradox for early-selection theories of attention? *Psychol. Sci.* **15**, 277–281 (2004)
10. Hubbard, E.M., Manohar, S., Ramachandran, V.S.: Contrast affects the strength of synesthetic colors. *Cortex.* **42**, 184–194 (2005)
11. Blake, R., Palmeri, T.J., Ma mis, R., Kim, C.-Y.: On the perceptual reality of synesthetic color. In: Robertson, L.C., Sagiv, N. (eds.) *Synesthesia: Perspectives from Cognitive Neuroscience*, pp. 47–73. Oxford University Press, Oxford (2005)
12. Paulesu, E., Harrison, J., Baron-Cohen, S., Watson, J.D.G., Goldstein, L., Heather, J., Frackowiak, R.S.J., Frith, C.D.: The physiology of coloured hearing: A PET activation study of colour-word synaesthesia. *Brain.* **118**, 661–676 (1995)
13. Segal, G.M.A.: Synaesthesia: implications for modularity of mind. In: Baron-Cohen, S., Harrison, J.E. (eds.) *Synaesthesia: Classic and Contemporary Readings*, pp. 211–223. Blackwell, Cambridge (1997)
14. Hanggi, J., Wotruba, D., Jäncke, L.: Globally altered structural brain network topology in grapheme-color synesthesia. *J. Neurosci.* **31**, 5816–5828 (2011). <https://doi.org/10.1523/JNEUROSCI.0964-10.2011>
15. Zamm, A., Schlaug, G., Eagleman, D.M., Loui, P.: Pathways to seeing music: enhanced structural connectivity in colored-music synesthesia. *NeuroImage.* **74**, 359–366 (2013). <https://doi.org/10.1016/j.neuroimage.2013.02.024>
16. Ward, J.: Synesthesia. *Annu. Rev. Psychol.* **64**, 4975 (2013)
17. Grossenbacher, P.G., Lovelace, C.T.: Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn. Sci.* **5**, 36–41 (2001)
18. Armel, K.C., Ramachandran, V.S.: Acquired synesthesia in retinitis pigmentosa. *Neurocase.* **5**, 293–296 (1999)
19. Brogaard, B.: Serotonergic hyperactivity as a potential factor in developmental, acquired and drug-induced synesthesia. *Front. Hum. Neurosci.* **7**, 657 (2013). <https://doi.org/10.3389/fnhum.2013.00657>
20. Shanon, B.: Ayahuasca visualizations: a structural typology. *J. Conscious. Stud.* **9**, 3–30 (2002)
21. Brogaard, B., Gatzia, D.E.: Psilocybin, LSD, mescaline and drug-induced synesthesia. In: Preedy, V.R. (ed.) *The Neuropathology of Drug Addictions and Substance Misuse*, vol. 2, pp. 890–905. Elsevier, New York (2016)
22. Myles, K.M., Dixon, M.J., Smilek, D., Merikle, P.M.: Seeing double: The role of meaning in alphanumeric-colour synaesthesia. *Brain Cogn.* **53**, 342–345 (2003)
23. Brogaard, B.: Synesthetic binding and the reactivation model of memory. In: Deroy, O. (ed.) *Sensory Blending: On Synaesthesia and Related Phenomena*, pp. 126–150. Oxford University Press, Oxford (2017)
24. Brogaard, B., Marlow, K., Rice, K.: The long-term potentiation model for grapheme-color binding in synesthesia. In: Bennett, D., Hill, C. (eds.) *Sensory Integration and the Unity of Consciousness*, pp. 37–72. MIT Press, Cambridge, MA (2014)
25. Smilek, D., Dixon, M.J., Cudahy, C., Merikle, P.M.: Synesthetic color experiences influence memory. *Psychol. Sci.* **13**(6), 548–552 (2002)
26. Brogaard, B., Vanni, S., Silvanto, J.: Seeing mathematics: perceptual experience and brain activity in acquired synesthesia. *Neurocase.* **19**(6), 1–10 (2012). <https://doi.org/10.1080/13554794.2012.701646>
27. Mills, C.B., Innis, J., Westendorf, T., Owsianiecki, L., McDonald, A.: Effect of a synesthete's photisms on name recall. *Cortex.* **42**, 155–163 (2006)
28. Bor, D., Billington, J., Baron-Cohen, S.: Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. *Neurocase.* **13**, 311–319 (2007)
29. Matey, J.: Can blue mean four? In: Bennett, D., Hill, C. (eds.) *Sensory Integration and the Unity of Consciousness*, pp. 151–170. MIT Press, Cambridge, MA (2014)
30. Rothen, N., Meier, B.: Higher prevalence of synesthesia in art students. *Perception.* **39**, 718–720 (2010)
31. Ward, J., Thompson-Lake, D., Ely, R., Kaminski, F.: Synesthesia, creativity, and art: what is the link? *Br. J. Psychol.* **99**, 127–141 (2008)

Color Syntax

► Color Combination

Color Term Translation

► Translation of Color Language

Color Terms

- [Color Categories and Vagueness](#)
- [Color Name Applications in Computer Vision](#)

Color Theory

- [Traditional Color Theory](#)

Color Transfer

Erik Reinhard
InterDigital, Cesson-Sévigné, France

Synonyms

[Color mapping](#)

Definition

Color transfer is the name given to a set of technologies that aim to alter the colors in an image or video by means of an example image or video. Thus, the mood or color scheme of an image can be changed to resemble that of another image, without the need to paint over pixels or manually alter an image's color palette.

Key Principles

Editing colors with traditional tools can be a laborious and time-consuming task. Professional color grading tools are complex and often expensive pieces of software that take time and effort to master. Rather than use buttons and sliders to adjust the color properties of an image, a convenient alternative may be to find an example image (also known as reference image) which already has the desired color mood. Using the colors

defined in the example image, color transfer methods try to alter the colors of the input image to match those of the example image.

The choice of example image is often determined by its use case. In a creative context the example image can be anything the user may imagine. When stitching panoramas, the individual shots may be matched to each other. In professional color grading, shot-to-shot matching may benefit from color transfer technologies help to unify the color schemes between shots. Here, the example image(s) will come from one of the shots.

Dependent on use case, different methods may be applicable. In the most general case, the input and example images are unrelated. In such case, statistical methods may be employed to transfer certain color characteristics from the example image to the input image.

If, on the other hand, it is known that the input and example images depict the same scene, then correspondences between the two images can be calculated, so that pixels depicting the same objects are known. This information can then be used to determine the mapping of colors between the two images in a meaningful and spatially varying manner.

The idea behind color transfer is that the user only needs to select an appropriate example image to apply its colors to an input image. This offers convenience, even if sometimes it may be difficult to predict the outcome. For this reason, non-automatic color transfer methods were developed that offer the user a high level of control over the result, without the need to resort to knobs and sliders to set the desired appearance.

The recent trend in computing to solve all manner of problems using deep neural networks has also given rise to network-based color transfer solutions.

Finally, color transfer is normally aimed to transfer the mood or appearance of an image to another image. However, the definition of mood or appearance is rather vague, and as a consequence there is typically no ground truth available. This makes it difficult to know when a particular transfer of color has been successful.

Each of these topics is discussed in the following sections.

Statistical Methods

Because human vision has three cones types which mediate color vision, color spaces can be three dimensional. Often this means that a color is specified as a mixture of red, green, and blue (R, G, and B). However, for natural images, RGB color spaces are highly correlated: the value of the red color channel tends to be a good predictor of the corresponding values of the green and blue channels. This is so because for natural scenes the average reflectance of its objects tends to gray. In an RGB color space this is represented as equal levels for the red, green, and blue channels. While correlations between channels can be used to create color constancy algorithms (see Color Constancy: https://doi.org/10.1007/978-3-642-27851-8_266-1), this is not good for easy manipulation of images.

One of the oldest color transfer algorithms [1] is based on a simple insight related to color spaces: if correlations between color channels can be minimized, then color transfer could be simplified to straightforward operations affecting each of the three color channels independently. For natural images, a well-chosen color opponent space can do just that: formally color opponent spaces can decorrelate the color channels. In practice, the channels become near independent [2].

Thus, by transforming both input and example images to such a color opponent space, the color channels can be processed independently. The simplest transfer of color properties then consists

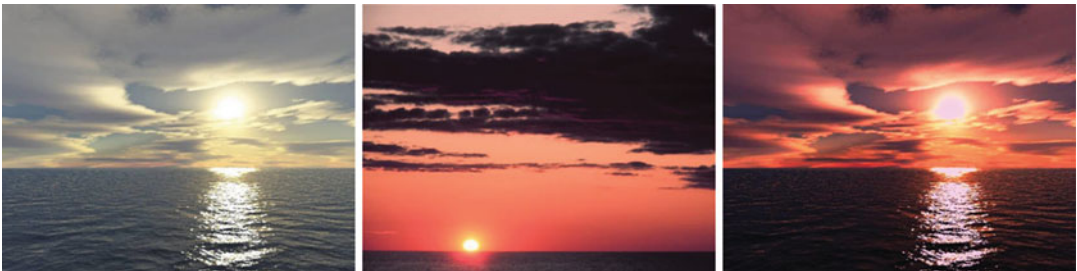
of shifting and scaling the pixels in each channel independently so that the input image will have the same overall means and standard deviations as the example image [1]. Then, the input image is transformed back to a suitable RGB color space. An example result obtained with this technique is shown in Fig. 1.

The assumptions of the above algorithm on the color statistics of the two input images are rather strong. It is assumed that the conversion to a color opponent space leads to 1D distributions that are independent from each other. They should each be unimodal. While this may be true approximately for some images, this would not be true in the general case. Especially if the target image has a single dominant color, for example mostly greens or reds, or the desired color palette is achromatic, then the transfer using the above method tends to produce reasonable results. This is largely the case for the example shown in Fig. 1. When the 3D distributions of the input and/or target images are more complex, the above method cannot match the target distribution, and other solutions would have to be sought.

Thus, such a simple algorithm can produce plausible results for a variety of input pairs, but equally there are many cases for which this approach fails. This has given rise to a great number of refinements, which each have contributed to either more robust or more plausible results.

For example, rather than relying on a single color space for all images, each input may be individually decorrelated [3], which can be achieved by applying Principal Components Analysis (PCA). However, a study on the choice

C



Color Transfer, Fig. 1 The two images on the left serve as input to a color transfer algorithm [1], to produce the result shown on the right

of color space has found that CIELAB using illuminant E provides the most plausible results, even out-performing PCA-based ensemble color spaces [2].

Alternatively, instead of simply transferring the means and standard deviations in each channel, a more sophisticated transfer may be achieved. For example, a partial and controllable histogram equalization could be applied in each channel independently [4]. It may also be constrained to avoid too large gradients in the output histograms [5].

Further, while using a decorrelated color space offers computational convenience, there are often corner cases which do not lead to plausible results. While on average per-channel methods may perform reasonably well, there will be many image pairs for which this is not the case. To increase the robustness of color transfer, it may therefore be advisable to consider applying a single transfer in 3D color space, instead of three one-dimensional and independent transfers [6]. This process may lead to some extraneous graininess, which can be removed using an appropriate post-process [7]. Examples of this method are shown in Fig. 2.

An interesting observation is that a good color transfer algorithm not only transfers the color palette from one image to another, but it also minimizes the amount by which each pixel is transformed. This can be achieved by formulating the problem as Monge's optimal transportation problem [8]. A linear optimal transport transfer is shown in Fig. 3. Non-convex relaxation may be applied as well [9, 10], as shown in the same figure.

Additional fidelity may be obtained by applying color transfer to image regions, rather than globally in one go to the entire image. To this end, the input and example images may be segmented or clustered, for example using expectation maximization to fit a Gaussian mixture model [11] or the mean shift algorithm. There are many other ways in which image regions can be clustered or the color transfer can be restricted or guided. This includes style-aware clustering, texture-based clustering, saliency-based color transfer as well as vectorization-based processing. Instead of transferring all colors of an example image to all pixels of an input image, a single



Color Transfer, Fig. 2 The two images at the top serve as input to an N-dimensional color transfer algorithm [6] (bottom left). Its output is subsequently passed through a regaining method [7] to produce the final image (bottom right)



Color Transfer, Fig. 3 The two images at the top of Fig. 2 serve as input to a color transfer algorithm based on Monge's optimal transportation problem [8] (left). The image on the right shows a result obtained using a different

Optimal Transport formulation [9, 10]. To enable a visual comparison, the bottom image shows Reinhard's method of color transfer [1]

object in an input image may be the target for color transfer.

Color space itself can also be clustered, for instance according to color naming conventions. This has the advantage that the transfer can be limited so that patches of a given name (e.g., "blue") stay within their color category.

Further fidelity can be obtained by also determining the material of objects in a scene. This knowledge could guide the transfer of colors, as demonstrated for portrait images, whereby image regions were classified as either skin, hair, or background [12].

After clustering, several techniques are available to perform the assignment of colors, including optimal transport theory [12].

Correspondence-Based Techniques

For applications whereby color adjustment is required to bring two or more images of the same scene in colorimetric alignment, the knowledge that the same scene is depicted may be exploited to either arrive at a simpler algorithm and/or at algorithms that operate with higher

fidelity. An example of a use case is panorama stitching, whereby multiple images of the same viewpoint are taken, which pairwise overlap. Here, color transfer may be used to ensure a consistent look between the images prior to stitching them together.

To aid the color transfer process, in such cases correspondences between the images may be computed first through feature detection and matching [13]. Points that correspond between two images can then be transferred, after designating one of the overlapping images as the input image and the other as the example image. Robustness of the transfer can be improved by considering neighborhoods around each feature point to derive the transfer function [14], for instance by using blur kernels or windowing.

Robustness could also be enhanced by explicitly using region-based feature matching techniques. Such an approach would typically involve either segmentation or a quad-tree decomposition.

In the case that one image of a pair needs to be corrected against another and the differences between them are rather large, the feature matching

algorithm should be chosen such as to produce a dense correspondence [15].

Non-Automatic Methods

While the paradigm to control an image's colors by example is attractive in many situations, there are times when more control over the final result may be required. In such a case, it may be possible to guide the transfer by a combination of an example image and additional user input.

Such user input may consist of strokes, whereby the user paints certain image areas. This can be used either to manually determine correspondences, or to prevent certain regions to participate in the transfer. Further, the user defined scribbles themselves may serve as initialization or reference to guide the color transfer. User input may also involve the selection of pairs of image regions, or swatches. These can be used to manually steer the color transfer.

Machine Learning Techniques

Machine learning techniques, and especially deep learning techniques, have enabled style transfer, whereby not only the colors are transferred between images, but the entire style is transferred [16]. This often involves example images of a painterly style, so that a photographic input image attains a non-photorealistic style. The transfer itself can be achieved using convolutional neural networks (CNNs) [16]. By introducing appropriate regularization and semantic segmentation to a CNN, style transfer can attain a photorealistic look, thereby becoming effectively a highly powerful color transfer technique [17].

Measuring Success

With much work devoted to improving the quality of color transfer, a question that has arisen is that of how to evaluate the quality of the results. In the absence of ground truth it is difficult to know when one algorithm produces a better result than another. Even defining "better" in this context is difficult, because it depends strongly on human perception, and even on personal taste. Measuring

the success of a color transfer algorithm is therefore a difficult problem.

However, some approaches may help with the assessment of color transfer methods. For instance, a robust color transfer algorithm would produce plausible results when input and example images are swapped. This can be tested in a meaningful manner by assessing the quality of a round-trip conversion. Here a color transfer algorithm takes an input and an example image to produce a result image. If then the same algorithm is used to transform the result image according to the input image, then in theory the final output should resemble the original input image [18]. After this round-trip conversion, a standard color difference metric such as CIE94 (https://doi.org/10.1007/978-3-642-27851-8_13-1) or CIEDE2000 (https://doi.org/10.1007/978-3-642-27851-8_7-1) could be applied to assess the per-pixel error, where a smaller error points toward a more robust algorithm.

Alternatively, a metric based on user data has been constructed [19]. Here, the color transfer quality was assessed in a psychophysical experiment. Using input, example and result images, a set of features are then computed on the same data, after which a regression method was applied to the combined features and user data. This allows the method to predict the quality of color transfer on unseen input/example/result triplets.

Cross-References

► [Color Constancy](#)

References

1. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graph. Appl.* **21**, 34–41 (2001)
2. Reinhard, E., Pouli, T.: Colour spaces for colour transfer. In: *IAPR Computational Color Imaging Workshop*. volume 6626 of *Lecture Notes in Computer Science*. Springer, pp. 1–15 (2011)
3. Abadpour, A., Kasaei, S.: An efficient PCA-based color transfer method. *J. Vis. Commun. Image Represent.* **18**, 15–34 (2007)

4. Pouli, T., Reinhard, E.: Progressive color transfer for images of arbitrary dynamic range. *Comput. Graph.* **35**(1), 67–80 (2011)
5. Xiao, X., Ma, L.: Gradient-preserving color transfer. *Comput. Graph. Forum.* **28**(7), 1879–1886 (2009)
6. Pitié, F., Kokaram, A., Dahyot, R.: N-dimensional probability density function transfer and its application to colour transfer. In: *ICCV '05: Proceedings of the 2005 IEEE International Conference on Computer Vision*, vol. volume 2, pp. 1434–1439. IEEE Computer Society, Washington, DC (2005)
7. Pitié, F., Kokaram, A., Dahyot, R.: Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.* **107**(2), 1434–1439 (2007)
8. Pitié, F., Kokaram, A.: The linear monge-kantorovitch linear colour mapping for example-based colour transfer. In: *4th European Conference on Visual Media Production*. (2007)
9. Rabin, J., Ferradans, S., Papadakis, N.: Adaptive color transfer with relaxed optimal transport. In: *IEEE International Conference on Image Processing*, pp. 4852–4856 (2014)
10. Rabin, J., Papadakis, N.: Non-convex relaxation of optimal transport for color transfer between images. In: *Geometric Science of Information*, pp. 87–95. Springer International Publishing (2015)
11. Tai, Y.W., Jia, J., Tang, C.K.: Soft color segmentation and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1520–1537 (2007)
12. Fu, Q., He, Y., Hou, F., Zhang, J., Zeng, A., Liu, Y.J.: Vectorization based color transfer for portrait images. *Comput. Aided Des.* **115**, 111–121 (2019)
13. Li, S.: A review of feature detection and match algorithms for localization and mapping. In: *IOP Conference Series: Materials Science and Engineering*, vol. volume 231, p. 012003 (2017)
14. Faridul, H.S., Stauder, J., Trémeau, A.: Optimization of sparse color correspondences for color mapping. In: *Color Imaging Conference*, pp. 128–134 (2012)
15. HaCohen, Y., Shechtman, E., Goldman, D.B., Lischinski, D.: Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph.* **30**(4), 70 (2011)
16. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423 (2016)
17. He, M., Liao, J., Chen, D., Yuan, L., Sander, P.V.: Progressive color transfer with dense semantic correspondences. *ACM Trans. Graph.* **38**(2), 13 (2019)
18. Faridul, H.S., Pouli, T., Chamaret, C., Stauder, J., Reinhard, E., Kuzovkin, D., Trémeau, A.: Colour mapping: a review of recent methods, extensions and applications. *Comput. Graph. Forum.* **35**(1), 59–88 (2016)
19. Hristova, H., Le Meur, O., Cozot, R., Bouatouch, K.: Perceptual metric for color transfer methods. In: *2017 IEEE International Conference on Image Processing*, pp. 1237–1241 (2017)

Color Trends

Maria Luisa Musso¹, Renata Pompas² and Leonhard Oberascher³

¹Architecture, Design and Urbanism, Universidad de Buenos Aires, Color Research Program, Buenos Aires, Argentina

²Milan, Italy

³FH Joanneum University of Applied Sciences, Salzburg, Austria

Synonyms

Color directions; Forecast; Temporal collective color preferences; Tendencies

Definition

Color trends are expected forthcoming market interest in some specific color shades.

Sociological Function of Color Trends

Color trends are temporal collective color preferences. They restrain individual- as well as group-specific color preferences and determine people's tastes and judgments over a certain period. Their application is typically not restricted to particular objects, contexts, or aesthetic conventions and is unconstrained by any functional, symbolic, or formal influences. Therefore, even colors with inherently negative symbolic meanings or connotations can become collectively preferred for a certain period. Individuals as well as society more generally perceive color trends as an articulation of "the predominant taste orientation." Anyone who adopts the trend demonstrates that she/he is open-minded, modern, and a member of an ideal (ized) peer group representing the *Zeitgeist*. In the end, using color trends also relieves individuals of mental strain. Those who adopt them avoid stylistic confrontation

and guarantee that their choices meet approval [1].

Cyclic Recurrence of Color Trends

Several studies [2–4], which analyzed the rise and fall of collective color preferences in architecture, interior design, and consumer goods during several decades, suggest a cyclic recurrence of color trends. Oberascher [5] points out that color trends must repeat in the long run because the potential gamut for color innovation is limited by the natural boundaries of the perceptual color space. New colors in a strict sense cannot be invented. Technological progress, however, may produce new color appearances. Based on an analysis of predominant color trends in furnishing and interior design in Germany and Austria between 1972 and 1992, he suggests a general model of the cyclic recurrence of collective color preferences. One explanation why certain color groups and ► **color combinations** are more likely to be repeated than others might be that the aesthetical evaluation and appreciation of color groups and combinations in general are determined by universal laws of perception and Gestalt psychology. But other factors may play a role. The emergence and spread of new collective color preferences are most probably rooted in a basic human desire (evolutionary, biological, neurophysiological, psychological, sociological) for change, alteration, renewal, and innovation. People's readiness to engage in new and unfamiliar colors increase over time as they become satiated with one trend. Since no new colors can be invented but only selected from the existing perceptual color gamut, those color groups and combinations that have been out of use for the longest time will appear novel. For young people these colors are new; for the older generations, they are a (re) contextualization of their color memories. The model of the cyclic recurrence of collective color preferences does not claim to predict color trends but might be used as a strategic tool for product marketing and management particularly in the furniture, furniture supply, and building material industry.

The Emergence of Color Trends

Historically, fashion has always set its rules and its trends based on a top-down pyramid approach. Every civilized society in its history has been able to use a large number of dye materials to color fibers, textiles, and leather in multiple shades. Where dye materials for some specific colors were lacking, international trade made up for it since they were imported, becoming widespread [6]. It can thus be said that no major color area has been neglected in the evolution of the trends and fashion of upper classes, but some color ranges have prevailed over others from time to time based on aesthetical, symbolic, and social choices imposed by the ruling classes.

A swing in trends occurred as ready-to-wear (prêt-à-porter) clothes and design became more widespread, and consumption democratization developed. Some spontaneous bottom-up trends have emerged since the 1960s. These have affected fashion, resulting in a combined top-down and bottom-up approach that is still underway, and it currently seems that street fashion is predominant [7].

Instruments to identify and anticipate market requirements were needed in order to properly respond to the new industrial organization of textile, clothing, furnishings, and design industries.

This has led to color trend forecast agencies, where teams of researchers (consisting of stylists, designers, sociologists, psychologists, and market experts) started to analyze consumption and behaviors and to anticipate color trends, ahead of the industrial manufacturing schedule required (2 years).

Not Only Colors, But Also Range Quality

Since then, color trends have renewed their offer season after season, not by changing major color families (red, yellow, green, etc.) but by the “quality of their range” (soft, bright, pastel, dusty, clear, dark, etc.). In the 1990s, surface finishing too became an element inherent to color as it changed its look, value, and “style.”

Historically, color trends have encapsulated the “spirit” of the time for many years. Their likelihood of meeting consumption demand results from the fact that they are based on the identification of color trends that are already found in the market, that the industry has adopted, and that the consumer has found within a predefined and limited available range, thus confirming relevant forecasts [8].

What Are Trends?

Trends are the expression of fashion themes and an indication of consumption and behavioral styles. They play two roles: on the one hand they are a monitoring tool for fashion and its multiple manifestations, and on the other hand, they anticipate and express current social orientations.

The life span of one or more trends on the market is very uncertain, from one season to several years.

The emergence of trends depends on historical, social, political, and cultural factors, which develop in the general public regularly and cyclically and influence almost a whole decade.

However, some very successful trends can come forth in a strange and sudden manner, as a result of occasional factors that have become very popular at a national or international level (a film, an exhibition, a music phenomenon, a product launch, the opening of new shops, some successful places open to the public, etc.).

Individualization of Color Trends

Today’s global market, where goods move freely, is unstable and difficult to control because many style orientations coexist simultaneously. Comparative analyses of major color trends often highlight dissimilarities in results, which are due to both the target and the impossibility to develop truly accurate forecasts. That is why color trends no longer play the role of universal consumption indicators; instead, they now act as a particularized anticipation of and focus on the themes intended for a specific market sector.

Companies use color trends to choose corporate colors for communication purposes as they try to develop a visible business identity, based on colors allowing them to stand out from competitors and to take advantage of mainstream tendencies by means of season’s colors. Albeit part of the general aesthetic framework, these color trends mark corporate choices and enable consumers to invent their own style through individual combinations.

This is the result of the current historical moment, known as “post-fashion,” where there are unclear and often contradictory signs and a fragmented coexistence of all types of different opportunities and orientations.

Creation of Color Trends

Color trends result from a research on ongoing social, cultural, and consumption changes, combining the quality and quantity data collected in order to develop an expected scenario.

The colors suggested by color trends have to play an intense and immediate communication role and must represent first a lifestyle than a consumption style one can identify oneself with. They must belong to the present time, be the continuation of the past, and prove to be able to anticipate the future.

As a matter of fact, the pursuit of novelties and changes cannot deviate much from tradition, from what consumers have shown to appreciate. Consequently, every color trend will contain some reassuring references in line with the latest successes and some innovations embodying the change.

Work Phases

Pompas points out the following work phases [9]:

- Registration of the latest most successful color trends
- Monitoring and insight of ongoing changes, based on a development process approach

- Collection of quantity and quality data on ongoing changes and their translation into dominant concepts
- Development of a scenario consisting of several images
- Selection of emerging shades and secondary shades
- Organization of a color range or ► [palette](#) that is consistent with dominant concepts
- Suggested ► [color matching](#), consisting in ► [color combinations](#)

Why to Be Aware of Trends

In the many areas of application and activities that involve the use of color, for instance, in industrial and textile design and in fashion design, only the creators of products who have the right information at the right time can act with advantages in the business environment.

There are many factors that need to be considered when predicting future design and color trends. There is an evolution from one season to the next; there are also important social and economic forces at work. One of the most important factors is the current and future projection of the socioeconomic conditions of the target consumer. Having a clear perception of changes, increasingly accelerated, it is possible to react quickly to the most demanding challenges of today's world.

The need arises from the spread now possible, at all levels, with simultaneity in all parts of the world, of the most important events. Due to the globalization of the marketplace, these factors are not as segregated by region as in the past. Of course, there are still major cultural influences that affect the interpretation of these factors.

Every new color or design trend starts because of a new influence or change of value perceived by the social group affected by these influences.

There are many reasons for the change of values, one of the most important being a shift in the emotional process of seeing and experiencing these changes. Visual and emotional influences, changes in economic and social

circumstances, and perception of new life styles affect the way in which people act in front of design and color. In a global environment of snapshot communications and accelerated pace of social and cultural changes, trends are constantly evolving to reflect those changes. Several organizations and color trend advisory services dedicate themselves to analyze the factors that will influence consumers' emotions. Their members track, analyze, predict, and direct colors and design of the consumer world. These groups base their color and design forecast on trends in fine arts, global events, technological advancements, economical and political circumstances, cultural facts, etc. and on the impact that these combined factors may have on the consumer. The media, opinion makers, specialized fairs, and producers deal with the effects of disclosing these facts and act on the choice of designs and colors. Based on this conceptualization, a new product will come, and it will be suitable to the market. If the designers have accurately identified and understood the customer, they will respond correctly to its own market.

Lifestyle Proposal

In 1985, the style of the products began to take into account the lifestyles [10].

The Advanced Communication Center, in France, divided the French population into five socio-lifestyles which then grew in number, diversifying. This study took into account:

- The places of residence of consumers
- Their behavior and preferences for the indoor environment: materials, objects, furniture, colors, and patterns
- The places where they buy or do their shopping
- The type of information used, newspapers, magazines, television and among other parameters

These data would compose an extremely useful tool to manage the offer of products.

Trends and Revivals

Particularly in the beginning of the twenty-first century, the succession of rapid changes in short and accelerated periods became relevant, leading to the overlapping of different revivals. The change in the nature of the geopolitical and cultural relationship impacts on the representation of the world, in the evolution of the conception and circulation of signs representing a moment. It is therefore essential the conceptual updating of the cultural and aesthetic debate, without neglecting the political, economic, and anthropological impact. It is of great interest therefore to assess how the arts of the image are carriers of elements of exchange, while it is necessary to avoid the danger of uniformity.

To propose a revival involves knowing the reality and the facts that gave rise and its implications. Taking only the outward, without understanding their reasons, leads to superficiality, the mere repetition. Investigating the roots pushes to motivation that is spilling into creativity.

Each of the decade 1970s, 1980s, and 1990s showed its spirit in a complex and multifarious use of color as a clear expression of the social premises of the time in western countries [11]. These trends are reflected in the product catalogs of European and American textile mills and department stores from 1970 to 2013, as well as in the main trade fairs such as Heimtextil (international trade fair for home textiles and commercially used textiles in Frankfurt), Star and Macef (Milan), Paritex and Maison et Objet (Paris), and among others, from 1971 to 2013.

The early 1970s kept the euphoric creativity of the 1960s as regards style. In 1973, however, great structural changes were brought about by the deep crude oil crisis which hit the markets and the economic structures. A feeling of uncertainty and lack of stability got hold of society. In the field of design, all this resulted in going back to a well-known past in search of the security purported in the everlasting values, either by retrieving old documents or manipulating eclectically their iconography. The desire of running away from reality was shown by all sorts of romantic attitudes such as revivals and retro-proposals.

Design turned back to small-scale motifs, geometrical or flowery. The *fleurettes* were monotonously added to all surfaces, and liberty style succeeded in west European countries. The predominant colors accompanying this stylistic shock ranged from medium to dark, grayish colors, non-saturated colors, earthy colors, grayish browns and greens.

The 1980s was the decade of appearance. The utmost egocentricity together with hedonism and obsession for social status was the more significant features in the group engaged in a blind and swift consumerism. It was the yuppie's decade, the decade of the self. The expanding social class was then the new bourgeoisie of the managerial elite, fond of showing off their success and their higher social position. The new style was sophisticated and luxurious, using sumptuous materials and emphasizing the quality in aesthetics. The essential feature was the pluralism, the multiplicity of styles, and the increasing variety of alternatives. The taste of the consumers was the result of their social status, keyed in their lifestyle and in the need of belonging to a social cultural group. It was a decade of a remarkable awareness of color, and thus color was conscientiously made the protagonist. High tech embraced industrial objects made of metal, glass, Formica, and plastic. Black made its appearance, stepping into the picture on its own, or associated with white and red, as well as with metallic colors, silver and gold, so as to emphasize the luxury effect [10, 12].

In the 1990s, the end of the Cold War brought about hope for the end of the nuclear menace. A new trend of thought set forward a deeper awareness of the environment in danger as much as greater concern about nature. Nature and its preservation became one of the main issues to deal with as well as a concern for health care and family life. A new conservative attitude was the most remarkable trend of this period, giving way to substantial subjects as ecology, the protective home, the enhancement of native roots and traditions, and the return to reassuring values. The new commandments were avoiding pollution, the efficient use of means and resources, the importance of quality instead of quantity, and the respect for nature. Attention was drawn in special toward

well-done high-quality work and craftsmanship. The search for true moral values was shown by the use of noble materials, and, in the choice of color, the main one was that of unbleached linen. The favorite colors were those of different types of wood, earth, grain, cereals, straw, sand, stone, and grayish colors that seemed to be worn out. Everyone celebrated nature in their own way: by means of choice of material, color, or subject [13, 14].

Trends for the House of the Century

The trends take into account the desire to expend more time at home, the valorization of empty space and customization of the environment, and the awareness for the preservation of natural resources. The market increasingly focused on products for the home. The decoration is freer, and nothing is definitive. Furniture is polyvalent, with fewer objects. The consumer seeks welfare, looks for quality, and wants a more simplified life. Pragmatism and health are at the order of the day.

In the new millennium, the house gains a new dimension. Tradition and modernity coexist in harmony. Earth, fire, and water are the three basic elements inspiring styles and colors. Individuality and daily rituals of life point the products chosen [14].

The configuration of a universal culture becomes a fundamental sign. The process of globalization tends to the unification of symbols, to the disappearance of diversity. It is therefore crucial to defend the values of each culture, using its imprint, to highlight its mark in every corner of the world, each time the triumph of cosmopolitanism can delete it.

From 2000 to 2020

The first 20 years of the new century saw the internationalization of markets, which diluted cultural and individual identities, and a radical change in languages and sensitivities. The contemporary context reshuffles all codes and erases any symbolic meaning. Colors thus lose their

psychological and emotional references in an operation of on-going recontextualization. All levels of consumption – the proposals of designers, fashion houses, brands, specialized magazines and the web – are increasingly subverted with an eclecticism that acts without suggestions from above and in which entropy has become the rule of color [15].

Cross-References

- [Appearance](#)
- [Color Combination](#)
- [Color Preference](#)
- [Dye](#)
- [Palette](#)
- [Pastel Colors](#)

References

1. Oberascher, L.: Colour trends – why we need them. In: Smith, D., Green-Armytage, P., Pope, M., Harkness, N. (eds.) *Proceedings of the 11th Congress of the International Colour Association (AIC 2009)*, CD. Colour Society of Australia, Sydney (2009)
2. Darmstadt, C.: *Farbbewegung in der Architekturgestaltung*. Unpublished manuscript (1985)
3. Koppelman, U., Kütke, E.: Präferenzwellen beim Gestaltungsmittel Farbe. *Marke. Z. Forsch. Praxis*. 2, 113–121 (1987)
4. Schlegel, M.: *Farb- und Materialscouting 1955–2005, Trends 2006–2007, Farbe – Struktur – Oberfläche*. Caparol Farben Lacke Bautenschutz (2006)
5. Oberascher, L.: Cyclic recurrence of collective colour preferences. In: Linton, H. (ed.) *Color Forecasting. A Survey of International Color Marketing*. Van Nostrand, New York (1994)
6. Pompas, R.: *The Trilogy of Precious Dyestuffs from the Mediterranean. Purple, Kermes, and Woad*. Textile Forum, Hannover (2001)
7. Luzzatto, L., Pompas, R.: *I colori del vestire. Variazioni–Ritorni–Persistenze*, Hoepli, Milan (1997)
8. Pompas, R.: La moda e il consumo del colore: quale approccio? In: Rizzo, S. (ed.) *Colore e design tra comunicazione e produzione*. De Ferrari, Genoa (2010)
9. Pompas, R.: *Textile Design. Ricerca–Elaborazione–Progetto*, Hoepli, Milan (1994)
10. Musso, M.L.: Black and status in the '80s. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) *AIC 2008, Colour – Effects & Affects, Proceedings of the Interim Meeting of the International Colour*

Association, Paper 119. Swedish Colour Centre Foundation, Stockholm (2008)

11. Musso, M.L.: Colour in the '70s, '80s, '90s, as expression of changes in society. In: Nieves, J.L., Hernández-Andrés, J. (eds.) AIC Colour 05, Proceedings of the 10th Congress of the International Colour Association, pp. 1589–1591. Comité Español del Color, Granada (2005)
12. Bayer, A.G.: Trend Colour Charts. Bayer AG, Leverkusen (1974–1990)
13. Musso, M.L.: Ecology and colour in the '90s. In: Ye, G., Xu, H. (eds.) AIC 2007, Color Science for Industry, Proceedings of the Midterm Meeting of the International Color Association, pp. 158–161. Color Association of China, Hangzhou (2007)
14. Hoechst, A.G., Trevira, GmbH: Trends in Living. Trevira Colour Chart. Hoechst & Trevira, Frankfurt (1981–2013)
15. Luzzatto, L., Pompas, R.: Il colore asimbolico. In: Colore e moda. Bompiani, Milan (2018)

Color Union

- [Color Combination](#)

Color Uses

- [Functionality of Color](#)

Color Vision Abnormality

- [Tritanopia](#)

Color Vision Screening

- [Color Vision Testing](#)

Color Vision Test

- [Pseudoisochromatic Plates](#)

Color Vision Testing

Galina V. Paramei¹ and David L. Bimler²

¹Department of Psychology, Liverpool Hope University, Liverpool, UK

²School of Psychology, Massey University, Palmerston North, New Zealand

Synonyms

[Clinical color vision tests](#); [Color vision screening](#); [Diagnosis of defective color vision](#)

Definition

Color vision testing is the assessment of chromatic discrimination ability and the diagnosis of any perceptual deficiency according to its severity and quality (► [Protanopia](#); ► [Deutanopia](#); ► [Tritanopia](#); Genetics of Color Vision). Tests vary in sensitivity, specificity, ease of use, and time required for administration [1–4]. Many were designed primarily for vocational screening for congenital deficiency, an issue in any occupation where color-coding conveys information (e.g., railways, aviation, electronics) [5]. Testing is also important for assessing and monitoring acquired color abnormality, appearing as a manifestation of visual system pathology resulting from ophthalmological diseases (e.g., glaucoma, ocular hypertension) [6], systemic or neurological diseases (e.g., diabetes [6], Parkinson's), or the effects of medications or exposure to environmental/occupational toxins (► [Color Perception and Environmentally Based Impairments](#)).

Color vision tests fall into several broad categories [1–4]. Pseudoisochromatic plates and arrangement tests both directly address an observer's confusions between color pairs along a given confusion axis. Both types of tests have the advantage of rapid administration and ease of interpretation, making them suitable for field/epidemiological applications. Generally, they distinguish between the protan, deutan, and tritan forms of deficiency. Their results emphasize a

dichotomous outcome: whether an observer's color sensitivity is (vocationally) impaired. Versions of the tests exist for testing color vision in children. More recent computerized developments measure variations in color perception along a continuous range. Matching tests and naming (lantern) tests follow different principles. Below, the most widely used tests are described in more detail.

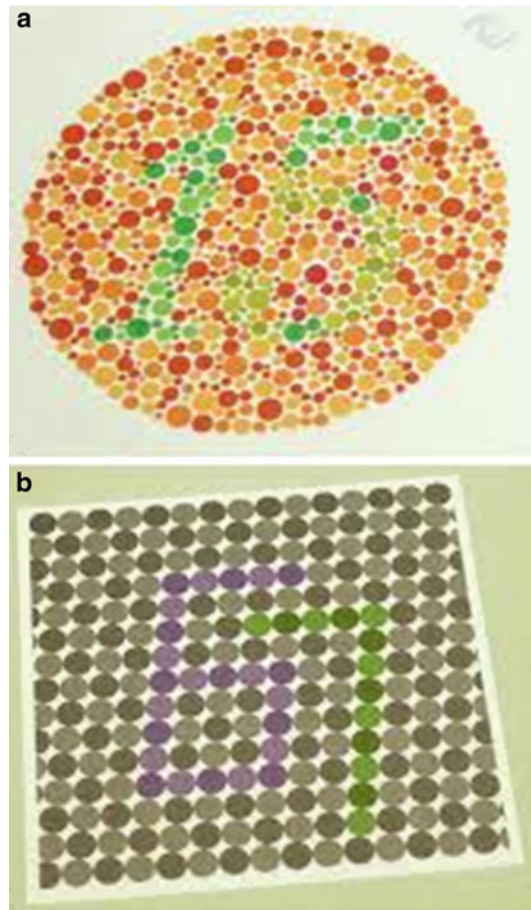
Pseudoisochromatic Tests

Pseudoisochromatic tests (Fig. 1) (► [Pseudoisochromatic Plates](#)) are irregular mosaics of small circles, randomly varying in size and luminance. Color differences among the circles demarcate a foreground design (digits, simple geometric forms, or curved lines), so that for a normal trichromatic observer the design stands out by Gestalt fusion from its background. In diagnostic plates the defining color difference disappears for color-deficient observers, and the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction. The spatial fluctuations in the mosaic mask any residual luminance traces of the normal design.

The Ishihara test, used most widely, is intended for diagnosis of congenital red-green deficiency ("daltonism"), differentiating its two types, protans and deutan, and severity (mild, moderate, or extreme) [2, 3, 8] (► [Protanopia](#); ► [Deuteranopia](#)). The Hardy-Rand-Rittler (HRR) test contains additional six plates designed to detect tritan defects and gauge their severity [8, 9]. There also exists a Farnsworth F2 plate designed specifically for revealing tritan abnormality (► [Tritanopia](#); ► [Color Perception and Environmentally Based Impairments](#)).

Color Arrangement Tests

Arrangement/panel tests use a set of color stimuli ("caps") which sample a color circle (► [Color Circle](#)) at regular intervals. The subject is requested to arrange them in sequence, so that



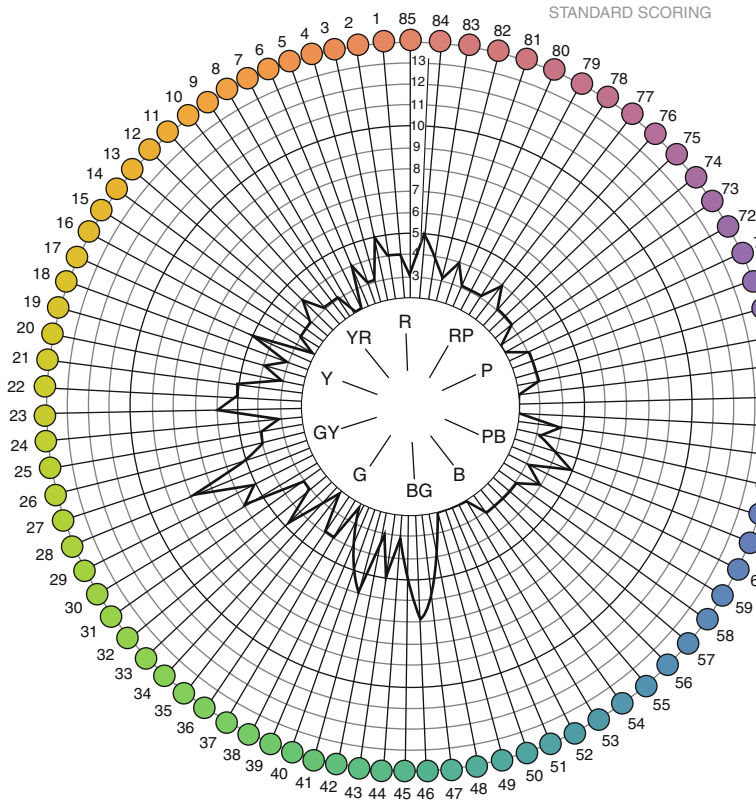
Color Vision Testing, Fig. 1 Examples of pseudoisochromatic plates used for screening for red-green deficiency (*top*) and tritan defects (*bottom*). (Source: Jägle et al. [7], Fig. 6.2. Springer Copyright Clearance Center. License Number: 3655300610668)

each color lies between the two colors most similar to it. Transpositions of the caps, departing from a normal trichromat's sequence, are recorded as errors. These departures can be plotted graphically to measure the angle of the confusion axis (if any) and summed to quantify the severity of any deficit.

The classical example is the Farnsworth-Munsell 100 hue (FM100) test [1–4, 9, 10], consisting of 85 caps, which takes 20–30 min to complete. Errors (transpositions) peak in the sectors of the color circle running tangential to the protan, deutan, or tritan confusion axis (Fig. 2). Performance on the FM100 improves with

Color Vision Testing,

Fig. 2 Example of the Farnsworth-Munsell 100 hue test scoring sheet indicating a moderate color discrimination defect. (Source: Jameson [11], Online Supplementary Material, p. 4, Fig. 3; <http://www.glimpsejournal.com/2.3-KAJ.html>. Copyright (2009) Kimberly A. Jameson, All Rights Reserved)



repetition, and as well as measuring color discrimination, it is affected by general nonverbal intelligence [12].

Two shorter versions, the Farnsworth Dichotomous D-15 test and the Lanthony desaturated D-15d, each contain only 15 movable caps plus a fixed “pilot cap” as the start of the sequence [10] and take about 5 min to complete. The D-15 is designed to diagnose moderate to severe color defects. The D-15d test uses color samples that are lighter and paler [13]. It was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. Errors can include diametrical circle-crossing transpositions, indicating the protan, deutan, or tritan confusion axis when plotted graphically (see Fig. 1 in ► [Color Perception and Environmentally Based Impairments](#)). Outcomes of both tests can be summarized as a color confusion index (CCI), where 1.0 corresponds to perfect color arrangement and CCI values greater than 1.0 indicate progressive impairment of color discrimination [2]. These tests are often used in

conjunction, though the more sensitive D-15d is widely employed for early detection of mild acquired dyschromatopsias.

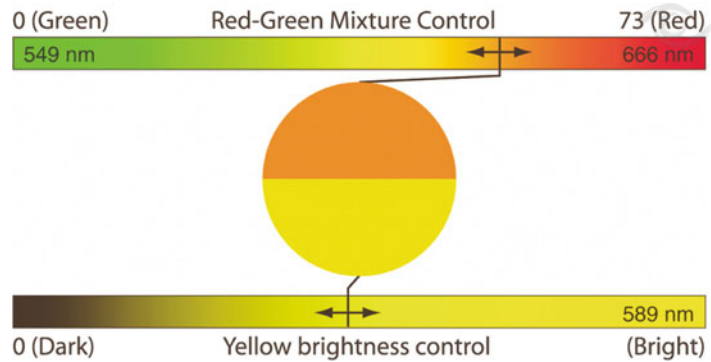
The Lanthony New Color Test [3, 4] comprises 4 panels of 15 caps each at 4 levels of saturation, to examine color similarities at 4 different scales. Like the D-15d, it can be used to track the progression of acquired dyschromatopsias.

Although not strictly an arrangement test, the City University Test [2–4] is derived from the D-15. It is a forced-choice test consisting of ten panels, each presenting four colored dots in a quincunx; the observer has to indicate which of the four colors most closely resembles the central one. The colors are selected so that protan, deutan, or tritan deficiencies affect which dot is subjectively most similar to the center.

Anomaloscopes

Within color-matching tests, the “gold standard” of color deficiency diagnosis is anomaloscopes,

Color Vision Testing,
Fig. 3 Rayleigh spectral
 matching in the Nagel
 anomaloscope. (Source:
 Jägle et al. [7], Fig. 6.5.
 Springer Copyright
 Clearance Center. License
 Number: 3655300610668)



which present colors as monochromatic lights rather than on a computer monitor or via reflective pigments. Compared to the pseudoisochromatic and panel tests, anomaloscopy requires a skilled examiner.

The Nagel anomaloscope is intended for assessment of red-green discrimination (Fig. 3). The observer views a 2° hemipartite circle, where one half is yellow light (589 nm) while the other half-circle mixes red (670 nm) and green light (545 nm) – known as the Rayleigh equation [1–3]. These wavelengths differ only in their relative stimulation of L- and M-cones (S-cones being unresponsive in this spectral range) (► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)). A normal trichromat's setting is characterized by a green/red ratio around 41 (on a scale between 0 and 73) and a very narrow range of such settings upon retest. Anomalous trichromats accept a wider than normal range of mixed lights as indistinguishable from the yellow, with the range – from narrow to very broad – characterizing mild, moderate, or extreme impairment. A dichromat can match *any* red/green ratio to the yellow light by adjusting the luminance of the latter, which a protanopic dichromat dims if the mixed light is mainly red whereas a green-dominated light requires a more luminant yellow; the opposite is observed for a deuteranope (► [Protanopia](#); ► [Deuteranopia](#)).

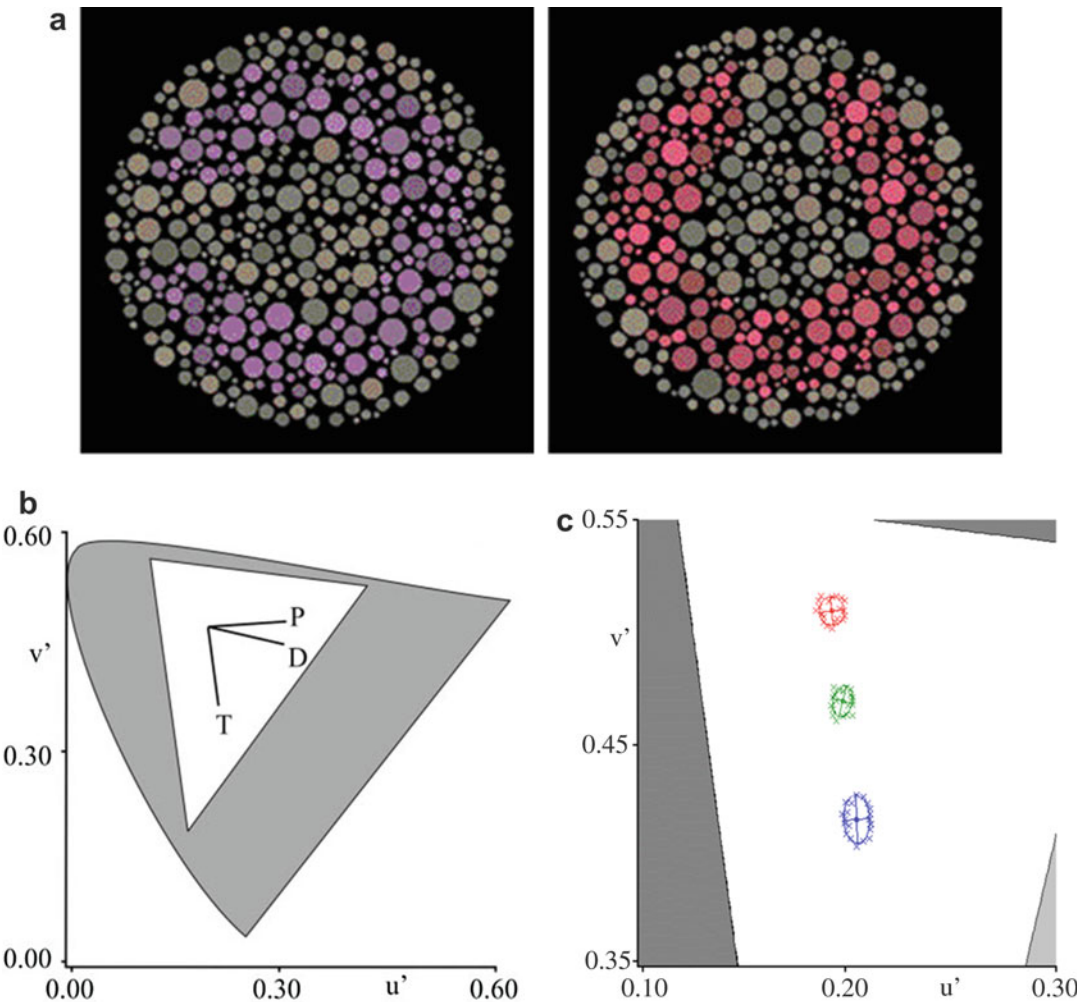
The Moreland anomaloscope serves to assess tritan discrimination. One half of the 2° hemipartite circle is a cyan standard (480 nm

light tinged with a small admixture of 580 nm), which must be matched by mixing indigo (436 nm) and green light (490 nm) in the other half-circle, known as the Moreland equation [14]. Decreasing discrimination along the tritanopic confusion lines increases the range of mixtures which perceptually match the standard. Notably, at 8° only complete tritanopes accept the full range of color mixtures as a match to the cyan standard (► [Tritanopia](#)).

Computerized Tests

More recently, computerized equivalents of pseudoisochromatic tests have become common, displaying a series of colored mosaics in which the elements vary in luminance spatially as well as dynamically to leave only chromatic cues. Employed on a calibrated monitor under strict psychophysical protocols, the tests allow precise measurement of chromatic sensitivity.

In the Cambridge Colour Test (CCT) [15] (Fig. 4a), the design in each display is a stylized letter “C” with a four-way choice for the orientation for the letter's open side (► [Protanopia](#); ► [Deuteranopia](#); ► [Tritanopia](#)). The magnitude of the color difference demarcating each design (i.e., the difficulty of the choice) varies interactively in response to the observer's ongoing performance, to specify the direction in the color plane of elevated thresholds and to “bracket” their discrimination in the frame of the CIE ($u'v'$)



Color Vision Testing, Fig. 4 The Cambridge Colour Test. (a) An illustration of chromatic targets, Landolt “C”, embedded in the luminance noise background. (b) Confusion vectors (in CIE 1976 $u'v'$ chromaticity diagram) along which the chromaticity is varied in the CCT Trivector test: protan (red), deutan (green), and tritan (blue). The origin of the vectors indicates chromaticity coordinates of the

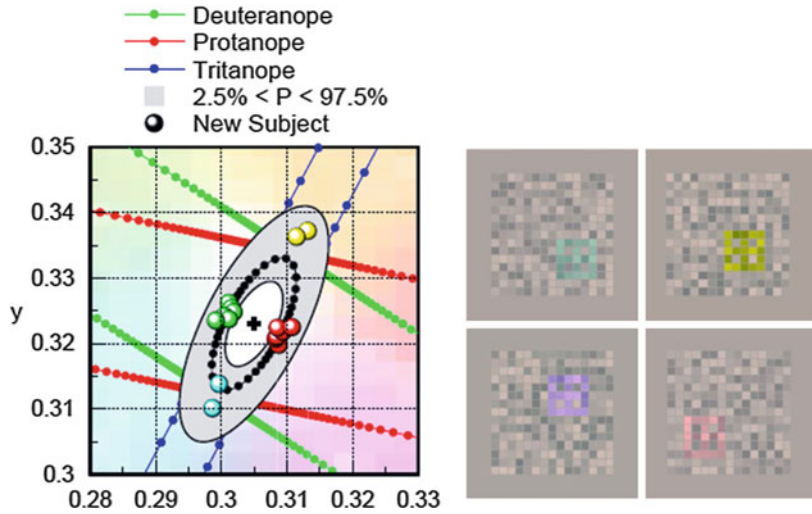
neutral background ($u' = 0.1977, v' = 0.4689$). (c) Examples of chromatic discrimination ellipses for normal trichromats: Ellipse 1 (middle), Ellipse 2 (top), and Ellipse 3 (bottom); crosses indicate raw discrimination vectors; fitted ellipses are shown by solid lines. (Fig. 4a, c: Source: Mollon and Regan [13], p. 4. Permission has been obtained from Prof. John D. Mollon, who holds the copyright)

1976 chromaticity diagram (► [CIE \$u', v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)). Outcomes are chromatic discrimination thresholds along the protan, deutan, and tritan confusion lines (Trivector test (Fig. 4b)), and elongation/orientation parameters of three MacAdam ellipses (Ellipses test (Fig. 4c)). CCT

normative data for normal trichromats for eight life decades track the impact of age upon chromatic discrimination [16]. The CCT has been employed in numerous clinical studies for differential diagnostics of relative damage to chromatic pathways [6].

Color Vision Testing,

Fig. 5 An illustration of the Colour Assessment and Diagnosis (CAD) test (right). Direction-specific, color-defined moving stimuli must be detected against a background with random, dynamic luminance contrast. (Source: Barbur and Rodríguez-Carmona [18], P. 52, Fig. 2.13(a, b). The authors hold the copyright; permission has been obtained from Prof. John Barbur)



The Colour Assessment and Diagnosis (CAD) test [5, 17] employs spatiotemporal luminance contrast masking. Direction-specific, color-defined moving stimuli must be detected against a background of random, dynamic luminance contrast (Fig. 5, right). Chromatic sensitivity is measured in 16 color directions in the CIE (x,y) 1931 chromaticity diagram. From these, mean thresholds are computed for the red-green and blue-yellow systems (Fig. 5, left) (Color Vision, Opponent Theory), to diagnose accurately the observer's class of color vision (i.e., normal, protan-, deutan-, tritan-like congenital loss, or acquired color vision deficiency). The CAD units of chromatic sensitivity are based on the mean thresholds measured in 333 young normal trichromats; they also serve as the reference for threshold estimates in normal aging [17]. The CAD test has been extensively applied for screening in occupations with visually demanding tasks that imply efficient chromatic discrimination [5] and in clinical practice [6].

“Electrodiagnosis” Tests

An alternative to behavioral testing is “electrodiagnosis” of congenital color vision deficiency, which measures retinal and cortical responses to stimulation, using standard electrophysiological

methods of electroretinogram (ERG) and visual evoked potential (VEP) recordings, respectively. Both forms of measurement show promise for detecting cone dysfunction. With a silent substitution paradigm to probe each cone type in isolation [19], L-cone stimulation evokes no ERG responses in protanopes (► [Protanopia](#)), while responses to M-cone stimuli are greatly reduced in deuteranopes (► [Deutanopia](#)); VEP amplitude is reduced in both classes of dichromats.

A second paradigm uses heterochromatic flicker stimulation, at high frequency (36 Hz) or intermediate frequency (12 Hz) to isolate the parasol-ganglion cell or midget cell retinal responses, respectively (i.e., luminance summation or chromatic opponency), while varying the counterphase red- and green-flicker contributions to overall stimulation [20]. The diagnostic parameters (luminance response fraction and L-cone response fraction) were found to reliably distinguish normal trichromats, protanopes, and deuteranopes.

Notably, in both paradigms, diagnoses have been confirmed by chromatic discrimination thresholds (measured with the Cambridge Colour Test) and genetic analysis of L- and M-opsin genes. The new diagnostic approach provides evidence of cone-specific ERGs and VEPs as reliable methods to detect congenital cone dysfunction.

Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Deuteranopia](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Tritanopia](#)

References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L.G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
3. Dain, S.: Clinical colour vision tests. *Clin. Exp. Optom.* **87**, 276–293 (2004)
4. Cole, B.L.: Assessment of inherited colour vision defects in clinical practice. *Clin. Exp. Optom.* **90**, 157–175 (2007)
5. Barbur, J.L., Rodríguez-Carmona, M.: Colour vision requirements in visually demanding occupations. *Br. Med. Bull.* **122**, 51–77 (2017)
6. Huchzermeyer, C., Kremers, J., Barbur, J.: Color vision in clinical practice. In: Kremers, J., Baraas, R. C., Marshall, N.J. (eds.) *Human Color Vision*, Ch. 10, pp. 269–315. Springer, Cham (2016)
7. Jägle, H., Zrenner, E., Krastel, H., Hart, W.: Dyschromatopsias associated with neuro-ophthalmic disease. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.) *Clinical Neuro-Ophthalmology. A Practical Guide*, Ch. 6, pp. 71–85. Springer, Berlin/Heidelberg (2007)
8. Birch, J.: Identification of red-green colour deficiency: sensitivity of the Ishihara and American Optical Company (Hard, Rand and Rittler) pseudo-isochromatic plates to identify slight anomalous trichromatism. *Ophthalm. Physiol. Opt.* **30**, 667–671 (2010)
9. Foote, K.G., Neitz, M., Neitz, J.: Comparison of the Richmond HRR 4th edition and Farnsworth-Munsell 100 Hue Test for quantitative assessment of tritan color deficiencies. *J. Opt. Soc. Am. A.* **31**, A186–A188 (2014)
10. Farnsworth, D.: Farnsworth-Munsell 100-Hue and dichotomous tests for color vision. *J. Opt. Soc. Am.* **33**, 568–578 (1943)
11. Jameson, K.A.: Human potential for tetrachromacy. *Glimpse J.: Art + Sci. Seeing.* **2**(3), 82–91 (2009)
12. Cranwell, M.B., Pearce, B., Loveridge, C., Hurlbert, A.: Performance on the Farnsworth-Munsell-100 Hue test is significantly related to non-verbal IQ. *Invest. Ophthalmol. Vis. Sci.* **56**, 3171–3178 (2015)
13. Lanthony, P.: The Desaturated Panel D-15. *Doc. Ophthalmol.* **46**, 185–189 (1978)
14. Moreland, J.D.: Moreland match revisited. *Vis. Neurosci.* **21**, 471–476 (2004)
15. Mollon, J.D., Regan, B.C.: *Cambridge Colour Test Handbook*. Cambridge Research Systems Ltd (2000). <http://www.crsLtd.com/tools-for-vision-science/measuring-visual-functions/cambridge-colour-test/>
16. Paramei, G.V., Oakley, B.: Variation of color discrimination across the life span. *J. Opt. Soc. Am. A.* **31**, A375–A384 (2014)
17. Barbur, J.L., Rodríguez-Carmona, M.: Color vision changes in normal aging. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) *Handbook of Color Psychology*, Ch. 8, pp. 180–196. Cambridge University Press, Cambridge (2015)
18. Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour vision: consequences for occupational environments. In: Best, J. (ed.) *Colour Design: Theories and Application*, Ch. 2, pp. 24–82. Woodhead Publishing Ltd, Cambridge (2012)
19. Barboni, M.T.S., Hauzman, E., Nagy, B.V., Martins, C.M.G., Aher, A.J., Tsai, T.I., Bonci, D.M.O., Ventura, D.F., Kremers, J.: Electrodiagnosis of dichromacy. *Vis. Res.* **158**, 135–145 (2019)
20. Aher, A.J., Martins, C.M.G., Barboni, M.T.S., Nagy, B.V., Hauzman, E., Bonci, D.M.O., Ventura, D.F., Kremers, J.: Electroretinographical determination of human color vision type. *J. Opt. Soc. Am. A.* **35**, B92–B99 (2018)

Color Wheel

- [Color Circle](#)

Color Word Translation

- [Translation of Color Language](#)

Colorant

- [Dye](#)
- [Pigment, Ceramic](#)

Colorant, Environmental Aspects

Gisela Umbuzeiro^{1,2} and Harold S. Freeman²

¹School of Technology, UNICAMP, Limeira, SP, Brazil

²Fiber and Polymer Science Program, North Carolina State University, Raleigh, NC, USA

Synonyms

Eco-/genotoxicity of synthetic dyes; Environmental chemistry of synthetic dyes; Structure property relationships

Definition

The environmental characteristics of colorants (dyes and pigments) encompass topics pertaining to the impact of such compounds on human health and biota. This is important because as much as 15% of the commercially used colorants reach the environment – mainly through the release of effluents into aquatic and soil biota. Humans can be exposed to various colorants or their transformation products often through drinking water and through dye and pigment manufacturing processes [1]. Therefore, it is important to have information pertaining to the toxicity of colorants as well as their possible transformation products, to avoid unacceptable health risks. In the case of dyes, the principal sources of losses are colorants remaining in dyebaths following textile dyeing processes. Another area of concern arises from the use of sludge from waste treatment plants as soil amendments, leading to possible soil contamination. Thus, methods pertaining to the treatment of industrial wastewater prior to the release of effluents have been developed, along with other pollution prevention methods, as key components of environmental stewardship.

Overview

The environmental properties of colorants are often determined by employing a battery of tests that are concerned principally with the potential for mutagenicity, carcinogenicity, and aquatic toxicity. An overview of developments in these areas is presented.

Mutagenicity

Some dyes are known to exhibit mutagenicity, and the most commonly used test for assessing mutagenicity potential is the reverse mutation assay employing specially engineered *Salmonella* bacteria. The standard *Salmonella*/mammalian microsome assay, often called the Ames test, is the most widely used protocol as an initial screening test procedure for assessing the mutagenic potential of new compounds. It was introduced in 1975 by Ames and co-workers [2], who observed that most mutagens were also carcinogens and that the extent of a compound's mutation of DNA was related to its carcinogenic potential, due to the susceptibility of DNA to chemical mutagenesis. Nowadays, the correlation between mutagenicity in the Ames test and carcinogenicity seems to be >60%.

The Ames test is an in vitro method that commonly uses one or more strains of *Salmonella typhimurium*. There are six strains of *Salmonella typhimurium* that are widely used in a mutagenicity test and are designed to detect different types of mutations involving colorants. TA98 and TA1538 are sensitive to frameshift mutagens, TA100 and TA1535 are used to detect base pair substitution mutations, and TA97 and TA1537 are used for frameshift mutations, which sometimes cannot be detected with the former strains. Also other strains such as YG1041, derived from TA98, have been employed because of high sensitivity to dyes containing nitro (NO₂) or amine (NH₂) groups. These strains cannot produce the amino acid histidine, an essential component for growth. Thus, the bacteria are unable to multiply unless a

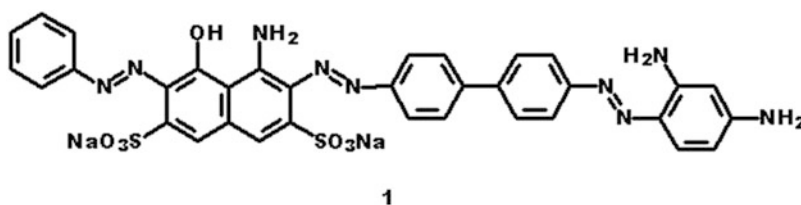
mutagen causes the proper type of reverse mutation in the histidine gene. Mutagenic activity can be measured quantitatively by simply counting the number of colonies present after incubating *Salmonella* bacteria with several doses of the test compound and other necessary test additives for a standard length of time.

The change in the bacteria that permits growth is called a reverse mutation, and the colonies that form are called revertant colonies. Generally, the test compound is considered mutagenic when the number of revertant colonies is more than twice that of the base count. Based on the number of the revertant colonies produced, the compound is characterized as a non-mutagen, a weak mutagen, or a strong mutagen. A solvent is added to facilitate adequate mixing, and an enzyme mixture may also be added to metabolize the test compound to enhance the sensitivity of the test, since bacteria do not have some of the enzymes present in mammalian tissues. The enzyme mixture is important

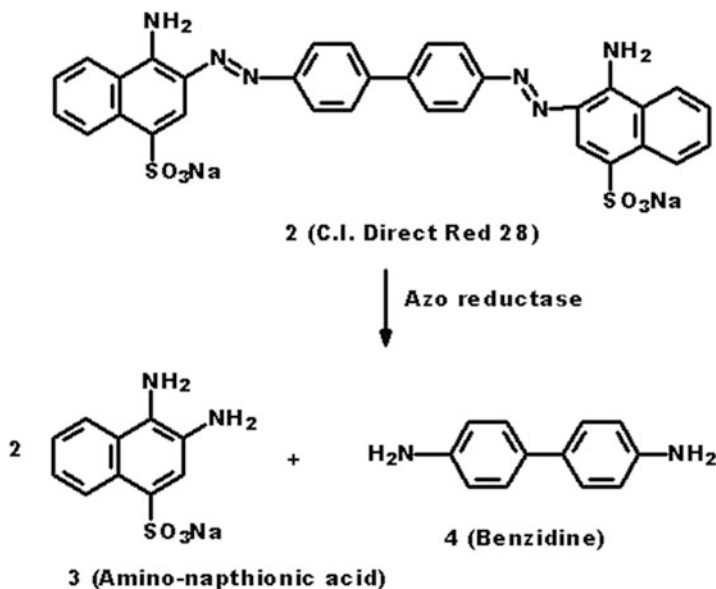
because many compounds are mutagenic once they are metabolized in the liver and other tissues. An exogenous metabolic system usually rat liver microsomal enzyme system, S9, is used in the *Salmonella* microsome mutagenicity assay.

Benzidine-based dyes such as C.I. Direct Black 38 (Fig. 1) are carcinogenic in several mammalian species. However, certain benzidine-based dyes (e.g., 1; C.I. Direct Red 28; Congo Red) are not mutagenic in the standard *Salmonella* microsome test. Consequently, the standard test was modified in 1982 to ensure the liberation of the parent diamine (cf., 2→4; Fig. 2) and the maximum possible mutagenic activity in each of the benzidine-based azo dyes studied [3]. This is an important protocol for benzidine-based dyes and is known as the preincubation assay (Prival modification) under reductive conditions, a test that employs hamster liver S9 which is believed to be richer in azoreductase enzymes.

Colorant, Environmental Aspects, Fig. 1 Molecular structure of C.I. Direct Black 38 (1)



Colorant, Environmental Aspects, Fig. 2 Reductase enzyme cleavage of C.I. Direct Red 28 liberating carcinogenic benzidine (4)



As a follow-up to positive Ames/Prival outcomes, *in vitro* testing such as a gene mutation assay involving mammalian cells or the mouse lymphoma test, a chromosome aberration assay, has been recommended [4]. This approach complements the mutagenic evaluation of a compound, providing a more accurate prediction of its mutagenic properties and carcinogenic potential.

Carcinogenicity

Although a variety of synthetic colorants have been studied, the majority of our knowledge and major concerns in this area arises from the recognition of azo dye manufacturing involving benzidine and 2-naphthylamine as a source of bladder cancer in humans. This soon led to extensive evaluation of azo dyes and their precursors for carcinogenic potential. The volume of results from work in this area became so substantial and diverse that a concerted effort to digest it has been made [5]. A key goal was to determine (1) whether the production of tumors in animal studies was sufficient to designate a dye as a human carcinogen and (2) the reliability of the published literature. This inspired an assessment of published data on a group of 97 representative azo colorants associated with the dyestuffs industry, for the purpose of determining the scientific standing of the experimental work in this area and the potential human hazard associated with these compounds. This assessment contributed to a set of recommended guidelines for carcinogenicity testing that cover chemical purity of test compounds, number of animals tested and survival rates, study controls, dose levels, route of administration, duration of experiments, pathological considerations, number of species, and evidence of human carcinogenicity. Application of the guidelines to published data afforded six categories of animal carcinogens and noncarcinogens:

1. Class A – the colorant was tested in at least two species of animals, resulting in the generation of nonconflicting data, using a sufficient number of animals (25 per group) and having a sufficient number of survivors for about two-thirds their expected lifetime. Repeated

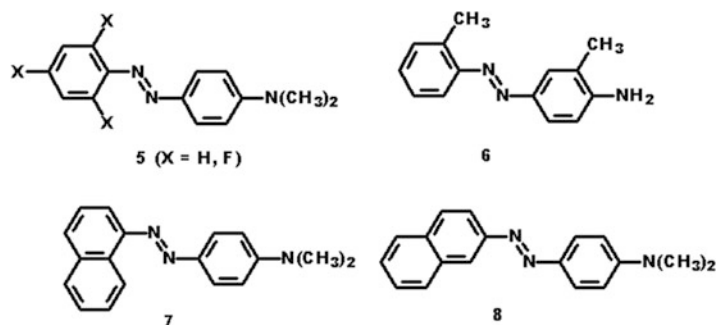
injections and urinary bladder implantations should not be regarded as meaningful in the context of chemically induced neoplasia. For a positive response, a target organ was identified and the induced tumors diagnosed as malignant.

2. Class B – the colorant was tested according to Class A guidelines, but meaningful data was generated for one species only.
3. Class C – testing of the colorant involved an insufficient number of test animals, and/or too many premature deaths occurred for results to be conclusive. For a positive result or trend, an increase in the number of animals with malignant tumors was observed but a target organ not identified.
4. Class D – the colorant was tested in multiple animal assays, which individually did not permit a conclusion but taken together provide sufficient evidence for an opinion.
5. Class E1 – the colorant was tested in a method considered inappropriate for evaluation of chemical carcinogenicity, e.g., by bladder implantation or using too few animals.
6. Class E2 – there was insufficient data to permit a rational judgment regarding carcinogenicity.

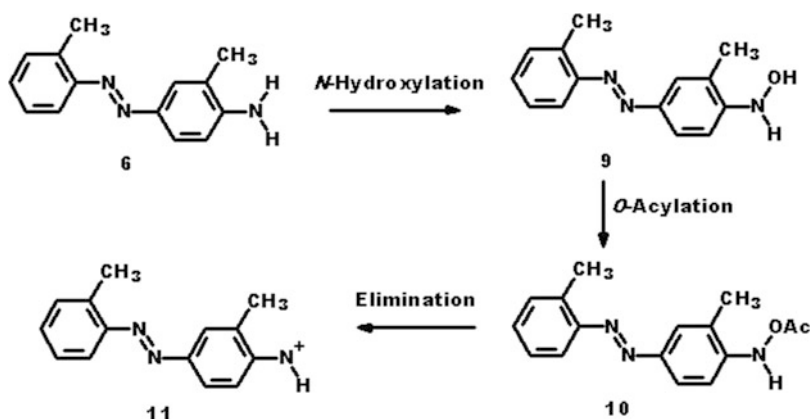
Structure-activity relationships associated with azo dye carcinogenicity produced the groupings depicted in Figs. 3, 4, 5, 6, 7, and 8 [6]. Group 1 comprises carcinogenic azo dyes that are hydrophobic (lipophilic) colorants having a 4-aminophenylazo base structure (5–8; Fig. 3) [7]. As illustrated in Fig. 4, the 4-amino group is susceptible to a series of metabolic steps leading to an electrophilic nitrenium ion (6→11) that can react with DNA. This illustration is important because it serves as a reminder that azo bond cleavage is not an absolute requirement for genotoxicity. Group 2 examples in Fig. 5 show that placement of the amino group ortho to the azo bond rather than para can lead to a nitrenium ion having an internal mechanism for deactivation (cf., 12→15). Group 3 comprises azo dyes producing one or more carcinogenic aromatic amines, upon cleavage of the azo bonds that are designated as carcinogenic, examples of which are shown in Fig. 6. In these examples, 2,4-dimethylaniline, 2,3,4-trimethylaniline, and

Colorant, Environmental Aspects, Fig. 3

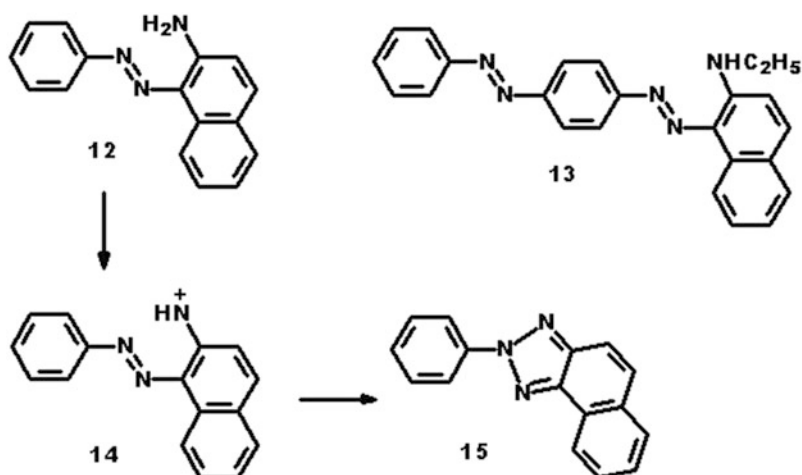
Carcinogenic
4-aminophenylazo
hydrophobic dyes in
Group 1

**Colorant, Environmental Aspects, Fig. 4**

Metabolism of a
4-aminophenylazo dye
leading to an electrophilic
species

**Colorant, Environmental Aspects, Fig. 5**

Noncarcinogenic
2-aminoazo dyes (**12–13**) in
Group 2 and metabolism
leading to triazole (**15**)
formation

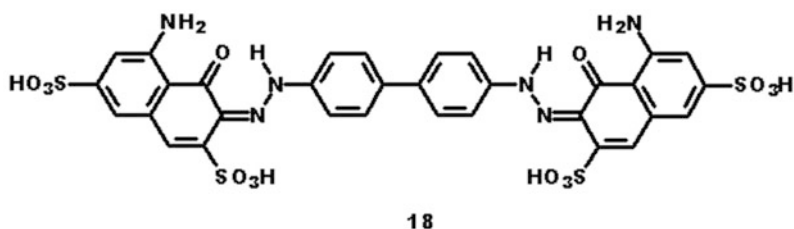
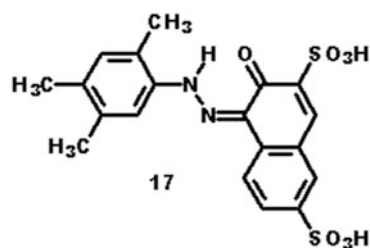
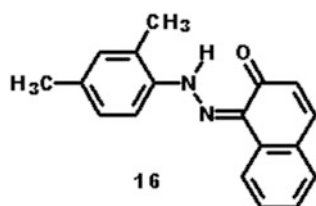


benzidine are produced by cleavage of the azo bonds in **16–18**. The properties of this dye grouping led to a ban on commercial products containing dyes derived from any of the family of 22 aromatic amines (now 24) known to be carcinogenic [8]. Similarly, dyes metabolized to

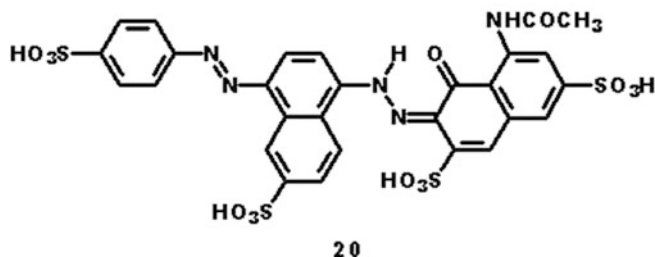
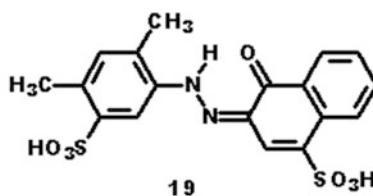
benzidine are now classified by the IARC (International Agency for Research on Cancer) as belonging to Group 1 (carcinogenic to humans) [9, 10].

Group 4 azo dyes are metabolized to noncarcinogenic sulfonated aromatic amines and are

Colorant, Environmental Aspects, Fig. 6 Azo dyes in Group 3 that liberate a carcinogenic aromatic amine upon reductive cleavage



Colorant, Environmental Aspects, Fig. 7 Azo dyes in Group 4 that liberate noncarcinogenic sulfonated amines upon reductive cleavage



regarded as safe to use (19–20). These structural types include FD&C Red 4 (19; Fig. 7), as well as FD&C Yellow 5 (Tartrazine, C.I. Food Yellow 4) and FD&C Yellow 6 (C.I. Food Yellow 3) [11]. Due to its benign nature, C.I. Food Black 2 (20) was the first black dye used for ink-jet printing and has long been a prototype for designing improved black dyes for this end-use area [12].

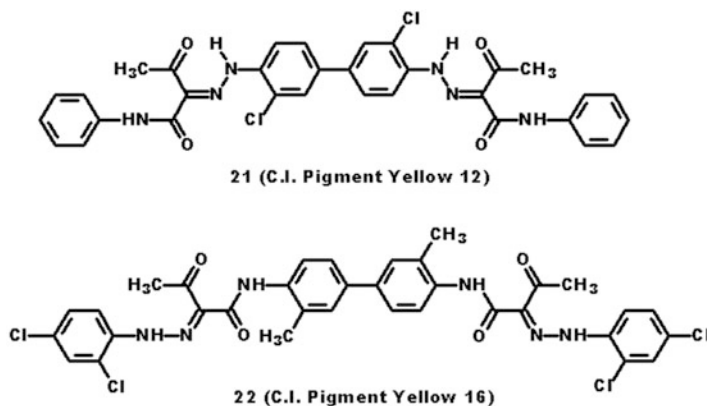
Insoluble organic pigments such as 21–22 characterize Group 5 colorants, which are non-carcinogenic (Fig. 8). They are not readily metabolized to their diamine precursors or absorbed into mammalian systems. The principal

requirement for safety among organic pigments is the control of the concentration of unreacted (free) aromatic amine precursors in the commercial product [13].

Aquatic Toxicity

Aquatic toxicology is the study of the deleterious effects of chemicals such as colorants on aquatic organisms. The associated tests may be used to measure the degree of response produced by exposure to various concentrations of dyes and pigments and may be conducted in the laboratory

Colorant, Environmental Aspects, Fig. 8 Examples of noncarcinogenic highly water-insoluble azo pigments in Group 5



or field. In the USA, the Clean Water Act (cf. Title 40 of the US Code of Federation Regulations 100–140, 400–470), which is administered by the Environmental Protection Agency (EPA), limits water pollution from industrial and public sources, stresses the importance of controlling toxic pollutants, and encourages investigations leading to waste treatment technologies [13]. Similarly, the Office of Pollution Prevention and Toxics (OPPT) implements the Toxic Substances Control Act (TSCA) as part of its responsibility for evaluating and assessing the impact of new chemicals and chemicals with new uses, to determine their potential risk to human health and the environment. Testing recommended includes acute and chronic toxicity for assessing environmental effects and bioconcentration, biodegradation, physicochemical properties, and transport/transformation studies for assessing environmental fate [14, 15].

Ecosystem-related toxicological testing includes acute and chronic toxicity tests employing species such as fish, *Daphnia*, algae, earthworms, and others. The overall goal is to determine the potential for changes in the composition of plant or animal life, reduction in reproductive success or the robustness of a species, alteration in the behavior or distribution of a species, and long-lasting or irreversible contamination of components of the physical environment (e.g., surface water). Specifically, acute toxicity tests are short-term tests designed to measure the effects of toxic agents on aquatic species during a short period of their life span. Acute toxicity tests

evaluate effects on survival over a 24- to 96-h period. Chronic toxicity tests are designed to measure the effects of toxicants to aquatic species over a significant portion of the organism's life cycle, typically one-tenth or more of the organism's lifetime. Chronic studies evaluate the sublethal effects of toxicants on reproduction, growth, and behavior due to physiological and biochemical disruptions.

The most common acute tests are performed with daphnids, mysid shrimp, amphipods, and fish (e.g., fathead minnow, zebrafish, and rainbow trout). Chronic tests are often used for life-cycle studies using *Ceriodaphnia dubia* (7 days) and *Daphnia magna* (21 days). LC50 (lethal concentrations 50%) are defined as the test concentrations at which 50% of the test organisms are dead after a certain time of exposure. For fish [16], some examples of LC50 values obtained for dyes were 0.047 mg/L for C.I. Basic Violet 1 (methyl violet), 0.12 mg/L for C.I. Basic Green 4 (malachite green), 1 mg/L for Disperse Blue 3, 3.2 mg/L for Basic Yellow 11, 4 mg/L for C.I. Basic Blue 3 and C.I. Acid Blue 113, 5.6 mg/L for C.I. Basic Brown 4, 5 mg/L for C.I. Mordant Black 11, 6.2 mg/L for C.I. Acid Green 25, and 7 mg/L for C.I. Acid Black 52. These commercial products included 29 of the 47 dyes with LC50 >180 mg/L, many that were bisazo or trisazo dyes. Three dyes had LC50 values between 100 and 180 mg/L, and the remaining 14 had values <100 mg/L. These results are consistent with the frequent lack of correlation between the genotoxicity and aquatic

toxicity of organic colorants. Whereas benzidine-based dyes such as C.I. Direct Black 38 (1) and C.I. Direct Blue 6 (18) are designated as carcinogenic, they did not exhibit acute toxicity when tested with aquatic organisms, along with their counterparts **23–24**, both of which are Cu complexes (Fig. 9).

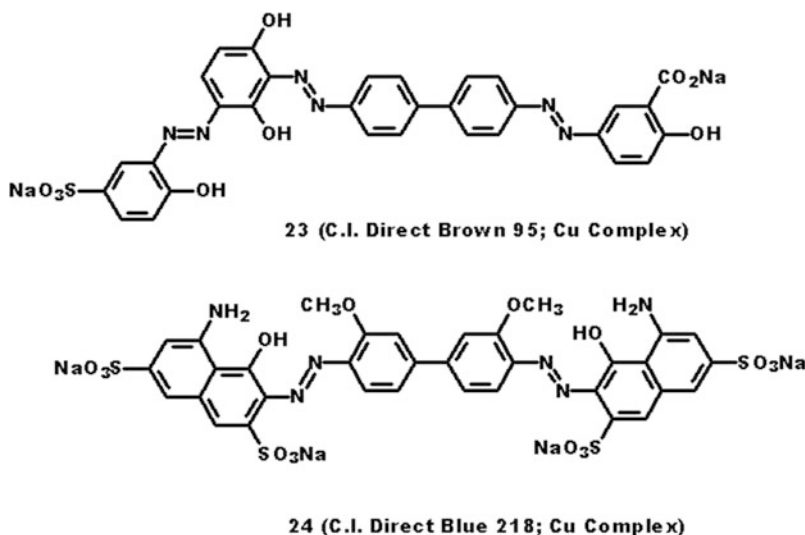
On the other hand, all of the cationic (basic) dyes (**25–28**) and anthraquinone disperse and acid dyes (**29–31**; Fig. 10) were significantly, if not extremely, toxic in this assay. Interestingly, bisazo dye C.I. Acid Blue 113 (**32**), unlike the bisazo direct dyes, unmetallized precursor C.I. Mordant Black 11 (**33**, Fig. 11), and its Cr-complex C.I. Acid Black 52 (**34**) were also significantly toxic. Water-insoluble vat dyes (**35–37**; Fig. 12) did not exhibit acute toxicity in this assay. Clearly, more studies are needed in this area in order to establish correlations between the molecular structures of colorants and their aquatic toxicity. While the picture is somewhat clearer regarding cationic dyes and vat dyes, it is much less so regarding sulfonated azo dyes. In a study involving the sparingly water-soluble monoazo dye C.I. Disperse Red 1, aquatic toxicity was observed in a variety of aquatic animals, and the microcrustacean *Daphnia similis* was the most sensitive species in acute and chronic toxicity tests [17]. In this study a risk assessment was performed. Acute and chronic water quality criteria were calculated

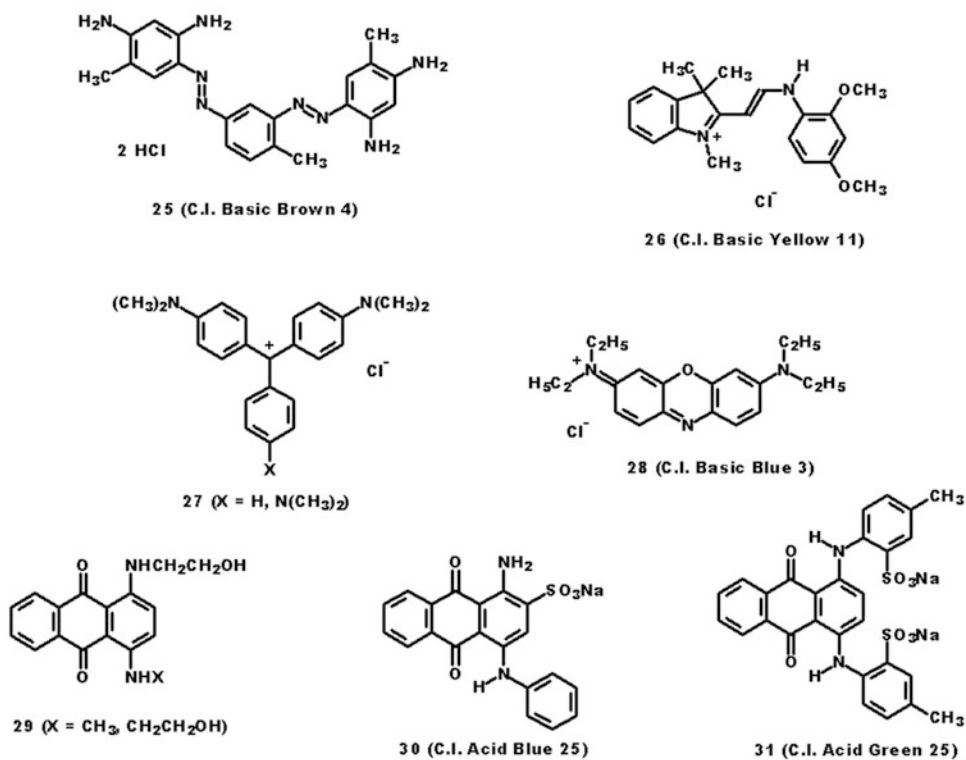
based on the toxicity tests. River water samples were analyzed during the 4 seasons of the year, and Disperse Red 1 was detected in 6 of the 14 samples. The concentrations varied from 50 ng/L to 500 ng/L. No acute risk was verified for any of the samples, but chronic risk was detected in five of the six samples [17]. Although dyes have been typically found in the aquatic environment and concentrations are in the ng/L range [18], it is important to have a comprehensive assessment of their hazard for a proper risk assessment evaluation.

Bioconcentration studies are performed to evaluate the potential for a chemical to accumulate in aquatic organisms that may subsequently be consumed by higher trophic-level organisms including humans. The extent to which a chemical is concentrated in tissue above the level in water is referred to as the bioconcentration factor (BCF). It is widely accepted that the octanol/water partition coefficient (e.g., Log K_{ow} or Log P) can be used to estimate the potential for nonionizable organic chemicals to bioconcentrate in aquatic organisms.

With regard to water quality concerns, a variety of methods have been developed to treat wastewater prior to industrial discharges, in order to remove colorants that are clearly visible at very low levels (e.g., 1 ppm). The methods employed involve chemical and biological agents for reduction and oxidation, physical adsorption,

Colorant, Environmental Aspects, Fig. 9 Examples of benzidine/benzidine congener-based dyes found negative in aquatic toxicity testing

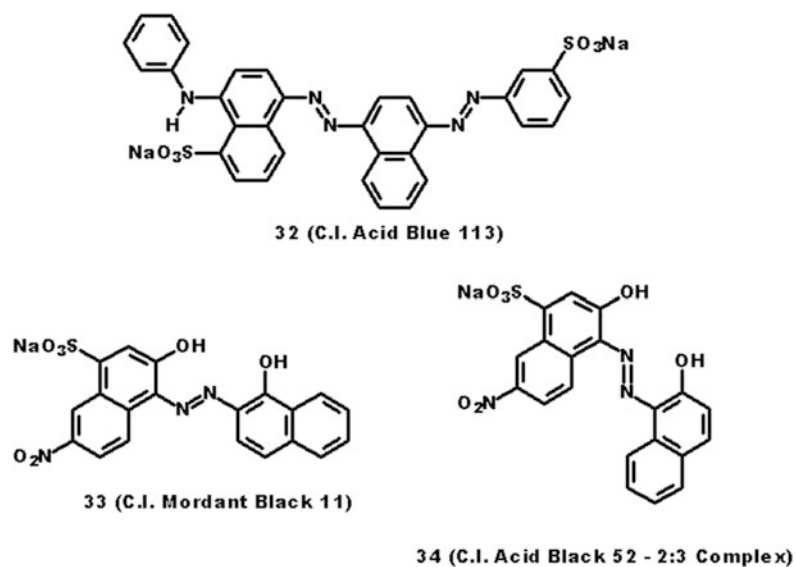


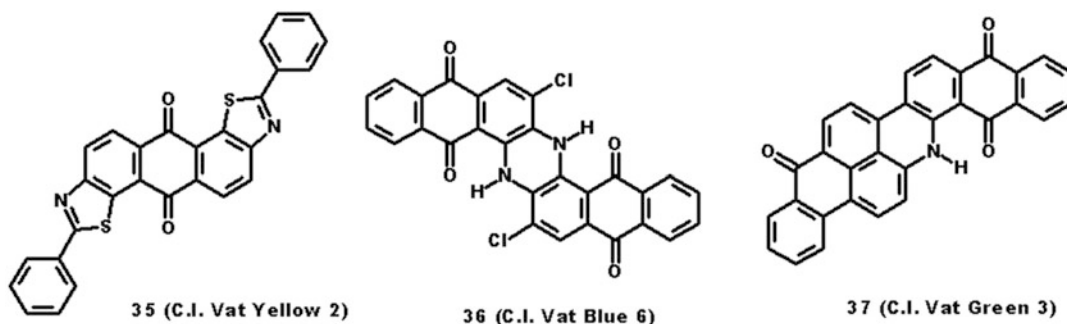


Colorant, Environmental Aspects, Fig. 10 Examples of cationic, disperse, and acid dyes exhibiting aquatic toxicity

Colorant, Environmental Aspects, Fig. 11

Examples of bisazo and metal-complexed acid dyes exhibiting aquatic toxicity

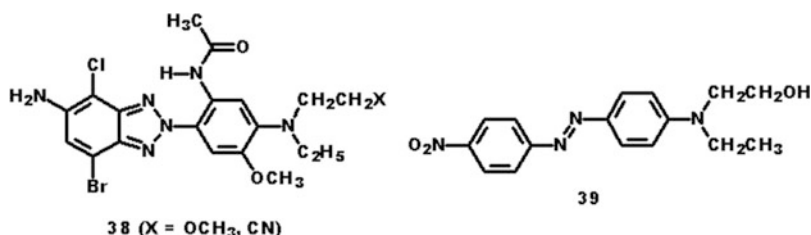




Colorant, Environmental Aspects, Fig. 12 Water-insoluble vat dyes found negative in aquatic toxicity testing

Colorant, Environmental Aspects, Fig. 13

Structures of compounds 38–39



membrane filtration, precipitation, and recycle and reuse. The latter methods are especially important since dyes are engineered to be stable compounds under end-use conditions and thus are completely decomposed with considerable difficulty. Also, care must be used when choosing a wastewater decoloration method. In several countries, effluents from biological treatment are decolorized using chlorine for simultaneous pathogen removal. It has been found, however, that the chlorination of wastewater containing certain disperse dyes can generate mutagenic compounds known as phenyl-benzotriazoles (PBTAs; **38**). These chlorinated compounds have been observed in drinking water. Similarly, the chlorination of commercial C.I. Disperse Red 1 (**39**; Fig. 13) produced higher mutagenicity than the untreated dye [19].

Future Directions

Bearing in mind that organic colorants include naturally occurring and synthetic compounds

and that the major concern generally lies with the latter types, there is renewed interest in exploring natural dye-based coloration processes because such dyes are widely regarded as biodegradable and of low inherent toxicity. In fact, various natural dyes are part of a normal diet (e.g., beta-carotene and lycopene, both of which are important antioxidants). Although few natural colorants have shown commercial viability (e.g., indigo, madder, logwood), they continue to find success in niche areas such as arts and crafts and enjoy exploration for potential use in mainstream products. It is worthwhile pointing out that the use of natural dyes as colorants for textiles often introduces the need for additives known as mordants, to help bond these dyes to fibers, since natural dyes have very low affinity for textiles. Mordants such as Cu and Cr are toxic to aquatic life and pose human health risks, making Al and Fe better choices. Regarding synthetic colorants, which encompass water- and/or solvent-soluble organic dyes and insoluble organic and inorganic pigments, they will continue to be the primary agents for adding color to substrates such as

natural and synthetic fibers, paper, plastics, petroleum products, printing inks, hair, and food, drug, and cosmetic products. Due to their practically insoluble nature, pigments have very low bioavailability; hence much of the attention in the area of environmental concerns has fallen on synthetic dyes. Among the synthetic dyes, azo dyes lie at the forefront in this area, and the choice of such dyes will continue to take into consideration the environmental properties of the required precursors and the metabolites formed in mammalian systems.

References

- Zollinger, H.: Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments, "Analysis, Ecology, and Toxicology of Colorants". Wiley-VCH, Cham (2003)
- Maron, D.M., Ames, B.N.: Revised methods for the Salmonella mutagenicity test. *Mutat. Res.* **113**, 173 (1983)
- Prival, M.J., Mitchell, V.D.: Analysis of a method for testing azo dyes for mutagenic activity in Salmonella typhimurium in the presence of flavin mononucleotide and hamster. *Mutat. Res.* **97**, 103 (1982)
- Hunger, K. (ed.): Industrial Dyes: Chemistry, Properties, and Applications, "Health and Safety Aspects". Wiley-VCH, Weinheim (2003)
- Longstaff, E.: An assessment and categorisation of the animal carcinogenicity data on selected dyestuffs and an extrapolation of those data to an evaluation of the relative carcinogenic risk to man. *Dyes Pigments.* **4**, 243–304 (1983)
- Gregory, P.: Azo dyes: structure-carcinogenicity relationships. *Dyes Pigments.* **7**, 45–56 (1986)
- Weisburger, E.K.: Cancer-causing chemicals. In: LaFond, R.E. (ed.) *Cancer-The Outlaw Cell*. American Chemical Society, Washington, DC
- German ban on use of certain azo compounds in some consumer goods. *Text. Chem. Color.* **28**(4), 11–13 (1996)
- IARC: Monographs on the Evaluation of Carcinogenic Risks to Humans. Some Aromatic Amines, Organic Dyes, and Related Exposures, vol. 99, 692 pp (2010)
- Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Coglian, V.: Some aromatic amines, organic dyes, and related exposures. *Lancet Oncol.* **9**(4), 322–323 (2008)
- Marmion, D.M.: Handbook of U.S. Colorants: Foods, Drugs, Cosmetics, and Medical Devices. Wiley, New York (1991)
- Carr, K.: Dyes for ink jet printing. In: Freeman, H.S., Peters, A.T. (eds.) *Colorants for Non-textile Applications*. Elsevier Science BV, Amsterdam (2000)
- Code of Federal Regulations, Title 40, Protection of Environment, US EPA, amended (1972)
- Fogle, H.C.: Regulatory toxicology: U.S. EPA/chemicals TSCA. In: Derelanko, M.J., Hollinger, M.-A. (eds.) *Handbook of Toxicology*. CRC Press, Boca Raton (1995)
- Adams, W.J., Rowland, C.D.: Aquatic toxicology test methods. In: Hoffman, D.J., Rattner, B.A., Burton Jr., G.A., Cairns Jr., J. (eds.) *Handbook of Ecotoxicology*. Lewis Publishers (CRC Press), Boca Raton (2003)
- Little, L.W., Lamb III, J.C.: Acute Toxicity of 46 Selected Dyes to the Fathead Minnow, *Pimephales promelas*. The University of North Carolina Wastewater Research Center, Department of Environmental Sciences and Engineering, Chapel Hill (1972)
- Vacchi, F.I., et al.: Occurrence and risk assessment of an azo dye -The case of Disperse Red 1. *Chemosphere.* **156**, 95–100 (2016)
- Zoccolo, G.J., et al.: Using SPE-LC-ESI-MS/MS analysis to assess disperse dyes in environmental water samples. *J. Chromatogr. Sci.* **53**(8), 1257–1264 (2015)
- Oguri, A., Shiozawa, T., Terao, Y., Nukaya, H., Yamashita, J., Ohe, T., Sawanishi, H., Katsuhara, T., Sugimura, T., Wakabayashi, K.: Identification of a 2-phenylbenzotriazole (PBTA)-type Mutagen, PBTA-2, in water from the Nishitakase River in Kyoto. *Chem. Res. Toxicol.* **11**, 1195–1200 (1998)

Colorant, Halochromic

► [Colorant, Thermochromic](#)

Colorant, Natural

Thomas Bechtold

Research Institute for Textile Chemistry and Textile Physics, Leopold Franzens University of Innsbruck, Dornbirn, Vorarlberg, Austria

Synonyms

[Natural colorants](#); [Natural dyes](#); [Pigments](#)

Definition

The term natural colorant comprises all kind of materials available from natural sources which are able to impart color to matter.

The main principle of color development for natural colorants is the specific absorption of light in the wavelength region of 400–700 nm [1].

Other principles on color formation may be based upon physical effects, for example, refraction of light (rainbows), interference (feathers of peacocks), or electron excitation (electroluminescence) [2].

Sources for natural colorants include minerals (red ochre, $\alpha\text{-Fe}_2\text{O}_3$), plant material (e.g., flavonoids from Canadian golden rod), and animal-based dyes (e.g., indigoid colorants from selected mollusk species).

Colored pigments are applied as finely divided solid particles which remain in an insoluble state during their application and use. Water-soluble or oil-soluble natural colorants can be found in food and beverage applications. A colorant that is used as dye must exhibit a specific substantivity towards that substrate. A dye is thus able to be sorbed at the surface of the material. Basic interactions lead to dyestuff sorption and binding on the material to be dyed are covalent binding, ionic attraction, π -bonding interaction, hydrogen bonding, and Van der Waals attractions [3]. With exception of covalent binding all other principles of dyestuff fixation are found in application of natural colorants.

According to structure, organic natural colorants can be divided into several main groups, namely, flavonoid dyes (including anthocyanins), substances containing naphthoquinoid and anthraquinoid groups, indigoid dyes, tannins, carotenoids, and chlorophylls [4].

The colorant content in natural sources is relatively low, usually of the order of a few % of the dry material. Thus, considerable amounts of raw material have to be processed.

Natural colorants are usually obtained by means of aqueous extraction of plant material; the use of solvents is less common as substantial amounts of plant material have to be extracted. As a result of the nonselective extraction process,

natural colorants usually contain a number of different color-providing chemical substances. The classification of a dye plant or animal source to a certain group of colorant usually is based on the properties of the most important species of dye present in the extract.

Major applications of natural colorants are as the coloring elements in food, as dyes for traditional textile dyeing eco-textiles, and as pigments for cosmetics [5].

Natural colorants can be applied in different ways. In many cases the extracted dye can be used directly (e.g., food and selected textile applications), but often mordants are used to increase dye adsorption (e.g., for textile dyeing and dye-lake formation). Important mordanting substances are tannins or metal salts. Due to ecological requirements mainly iron and aluminum salts are used as metal-based mordants.

Major Classes of Natural Colorants

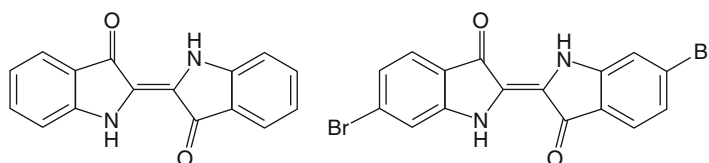
As mentioned, as a result of the extraction of natural colorants from native sources, with few exceptions, a number of colored substances are present in the extract, and, therefore, natural colorants are classified on the basis of the main constituent, which is of relevance for a given application. Important dyes with known chemical structure have been assigned color index (C.I.) generic names and/or C.I. constitution numbers. The major classes of colorant are discussed below.

Indigoid Dyes

The two most important natural dyes of indigoid structure are indigo (C.I. Natural Blue 1) obtained from plant sources and Tyrian purple (6,6'-dibromoindigo, C.I. 75800) from the Muricidae family, e.g., spiny dye murex (*Bolinus brandaris*) or banded dye murex (*Hexaplex trunculus*).

The structures of these two dyes are given in Fig. 1.

Tyrian purple is found in extracts from the hypobranchial glands of the gastropod mollusk. While the applicatory properties and fastness of the colorant were generally acceptable, the use of

Colorant, Natural,**Fig. 1** Structure of indigo (left) and Tyrian purple (right)**Colorant, Natural, Table 1** Indigo plants

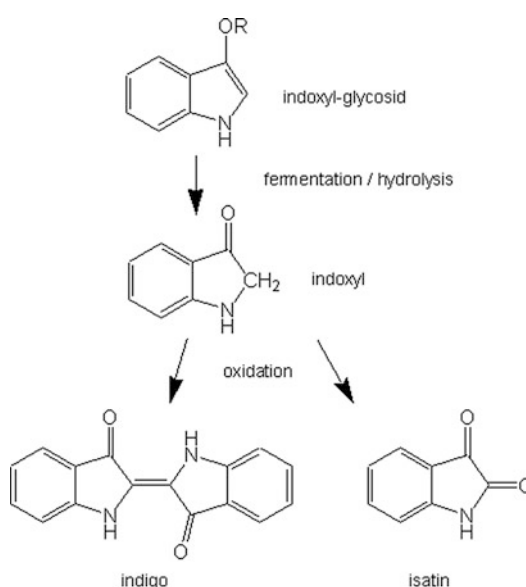
Plant	Botanical name	Climate	Region	Precursor
Indigo plant	<i>Indigofera tinctoria</i> L.	Tropical	India, Africa, and North, Central, and South America	Indican indigo- β -d-glucoside
Woad	<i>Isatis tinctoria</i> <i>Isatis indigotica</i>	Temperate	Europe, China	Isatan B indoxyl-5-ketogluconat
Dyer's knotweed, Ai	<i>Polygonum tinctorium</i> AIT	Subtropical and temperate	Europe, Japan	Indican indigo- β -d-glucoside

this dye ceased due to the low dye content in the mollusk.

Indigo is the most relevant blue natural dye and different plants have been cultivated for indigo production all over the world. As a general rule, the indigo is present in the plant in the form of different colorless indoxyl glycosides as precursors. Selected plants for indigo production are summarized in Table 1. The indigo precursor indoxyl also can be synthesized by microbiological methods [6].

To obtain indigo the precursor is first hydrolyzed enzymatically to obtain the intermediate indoxyl, which then rapidly oxidizes in the presence of air oxygen to indigo. The precipitated solid indigo dye then can be collected. The oxidation of indoxyl to yield the yellow isatin is an unwanted side reaction, which can cause substantial losses in indigo yield. A general scheme is presented in Fig. 2.

Besides the use as insoluble indigo colorant, indigoid dyes usually are applied as vat dyes. In vat dyeing the insoluble dye (indigo) is firstly converted to the alkali-soluble leuco form, which exhibits substantial substantivity towards protein fibers (wool and silk) and cellulosic fibers (cotton, flax, and regenerated cellulose fibers). The adsorbed reduced indigo dye (yellow) is then oxidized to generate the blue-colored insoluble form of the colorant using air or chemical oxidants. For deep shades, the procedure of immersion into the

**Colorant, Natural, Fig. 2** Main steps in production of natural indigo (R = glycosidic bond carbohydrate)

leuco dye bath and subsequent oxidation in air is repeated several times.

Dye reduction can be achieved by use of reductants and alkali, such as sodium dithionite ($\text{Na}_2\text{S}_2\text{O}_4$) or reducing sugars and NaOH ; alternatively, anaerobic microbial reduction can be employed. Microbiological dye reduction is practiced in traditional handicraft dyeing, but up to now is not applicable for dyeing with indigo in denim production (jeans).

Flavonoids and Anthocyanins

Flavonoids contribute to the color of many food products and thus are consumed by humans in the form of fruits and food. The color gamut provided by these colorants ranges from the yellow flavonoids (e.g., flavonols, chalcones, and aurons) to orange–red–purple anthocyanins. In plants many flavonoids are present as glycosides; anthocyanins are also glycosides, while the respective aglycones are named anthocyanidins. Representative structures of flavonoids and anthocyanidins are given in Fig. 3.

Yellow flavonoid dyes can be extracted from a high number of plant sources; selected representatives are given in Table 2 [7].

For textile dyeing, flavonoids can be used as direct dyes or as mordant dyes in combination with metallic mordants such as Fe or Al salts. When Fe, Sn, or Cu salts are employed as mordants, metal complexes are formed and a distinct change in color is observed, e.g., olive shades are obtained when an onion extract is applied in combination with Fe mordant.

The range of colors of anthocyanin dyes is of high coloristic interest for application as food colorants and for general purposes of coloration. The limited stability of the molecules can be improved by co-pigmentation, metal complexation, and the use of additives. Dependent on solution pH, anthocyanins rearrange to form different

species with different absorption spectra. Some of these products are colorless which explains the propensity of the colorants to fade.

A general scheme describing the pH-dependent rearrangement of an anthocyanin structure is given in Fig. 4.

At pH 1.0, orange–red/violet oxonium/flavylium ions are present which are of greatest stability (Fig. 4, AH⁺). With increase in pH to between 4 and 6, the colored quinoid base appears (Fig. 4, A). At the same time the formation of the colorless hemiketal (Fig. 4, B) gains importance. The hemiketal B is sensitive to tautomerization and ring opening, thereby leading to colorless cis- and trans-chalcone forms (Fig. 4, Cc and Ct) [8].

Stabilization of the dye molecules can be achieved by co-pigmentation. By inter- and intramolecular complex formations, self-association, and metal complex formation, the stability of the dyes can be increased; in many cases the color becomes bluer (bathochromic shift) and absorbance is increased (hyperchromic shift).

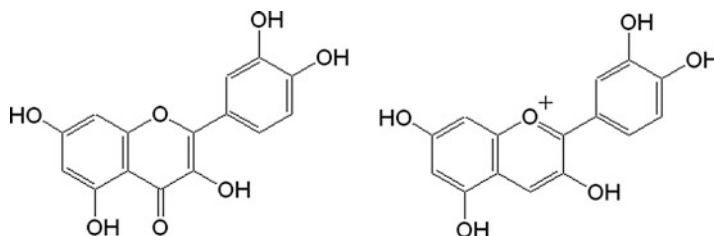
Anthocyanins are found in many intensively colored fruits and berries. Examples are summarized in Table 3.

Quinoid, Naphthoquinoid, and Anthraquinoid Dyes

Important yellow, orange, and red dyes belong to this group of colorants.

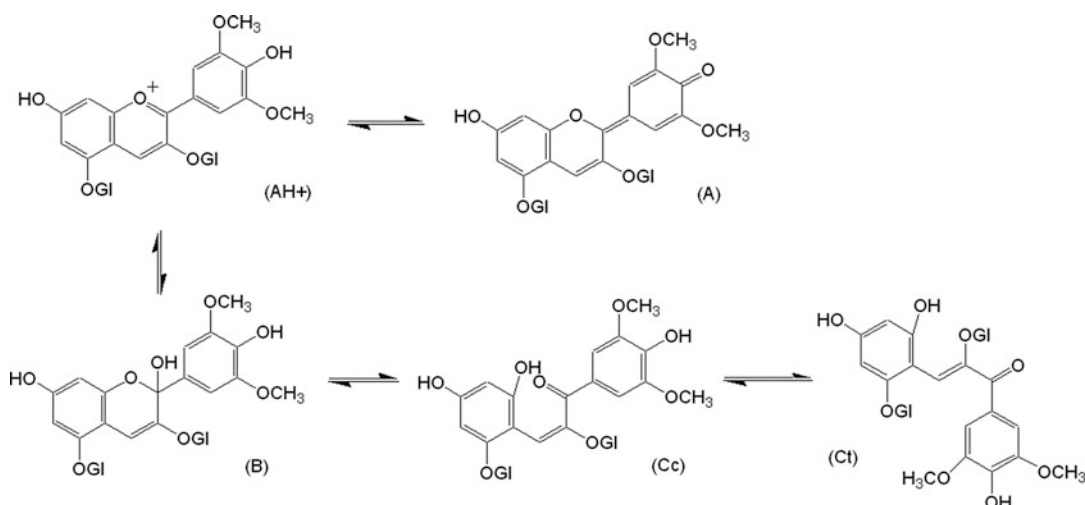
Colorant, Natural,

Fig. 3 Basic structures of quercetin (flavonoid, left) and cyanidin (anthocyanidin, right)



Colorant, Natural, Table 2 Plant sources for flavonoids

Plant	Botanical name	Main colorants	Color index number	Part of plant
Weld	<i>Reseda luteola</i>	Luteolin	C.I. Natural Yellow 2	Plant except roots
		Apigenin		
Roman chamomile	<i>Chamaemelum nobile</i>	Apigenin	C.I. Natural Yellow 1	Flower
Onion	<i>Allium cepa</i>	Quercetin		Outer shell of fruit
Black oak	<i>Quercus velutina</i>	Quercetin	C.I. Natural Yellow 10	Bark

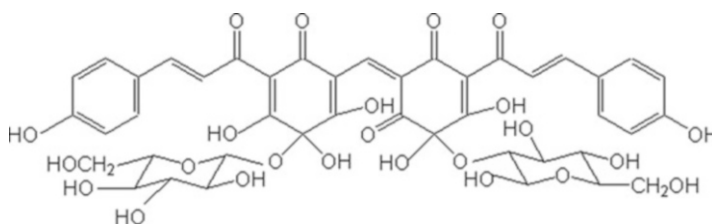


Colorant, Natural, Fig. 4 pH-dependent structural transformation of anthocyanins (malvidin 3,5-diglycoside)

Colorant, Natural, Table 3 Plant sources for anthocyanins

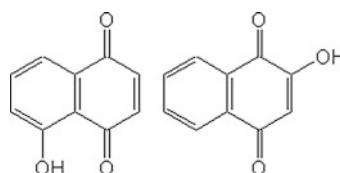
Plant	Botanical name	Main colorants	Color index number	Part of plant
Elder	<i>Sambucus nigra</i>	Cyanidin glycosides		Fruit
Vine	<i>Vitis vinifera</i>	Malvidin glycosides		Fruit
Privet	<i>Ligustrum vulgare</i>	Malvidin, cyanidin, delphinidin glycosides	C.I. Natural Black 5	Berries
Hollyhock	<i>Alcea rosea</i>	Malvidin, delphinidin glycosides		Flowers

Colorant, Natural, Fig. 5 Structure of carthamin



The red–orange quinoid dye carthamin (C.I. Natural Red 26) can be obtained by extraction of blooms of safflower (*Carthamus tinctorius*) (Fig. 5).

Two important representatives of naphthoquinone dyes are lawson (2-hydroxy-1,4-naphthoquinone) and juglon (5-hydroxy-1,4-naphthoquinone) (Fig. 6).



Colorant, Natural, Fig. 6 Structures of lawson (left) and juglon (right)

Lawson (C.I. Natural Orange 6) is extracted from henna (*Lawsonia inermis*). The plant grows best in tropical savannah and arid zones. The leaves contain lawson in the form of its glycoside, which hydrolyses and releases the aglycone, which then is oxidized to the quinone form. Lawson is of significant interest in hair dyeing. As a small molecule, its diffusion rate is sufficiently high to impart acceptable levels of color under the gentle conditions employed for hair dyeing. When henna leaves are mixed with leaves of the indigo plant, the so-called black henna is obtained. In the application as hair dye, lawson and indigo are fixed on the hair and brown-black shades are obtained.

Anthraquinone dyes can be extracted from plant sources (madder, *Rubia tinctorum*, C.I. Natural Red 8; *Hedge bedstraw*, *Gallium mollugo*, Natural Red 14) [9] or from animals (Kermes, *Kermes vermilio*, C.I. Natural Red 3; cochineal, *Dactylopius coccus*, C.I. Natural Red 4). Anthraquinoid dyes also have been extracted from fungi and lichen [10].

Most plant-based anthraquinoid dyes are extracted from the roots of the plants. The colorants are present as glycosides which hydrolyze during storage or extraction to the corresponding anthraquinoid dye. Representative structures (Alizarin, 1,2-dihydroxyanthraquinone, pseudopurpurin 3-carboxy-1,2,4-trihydroxyanthraquinone)

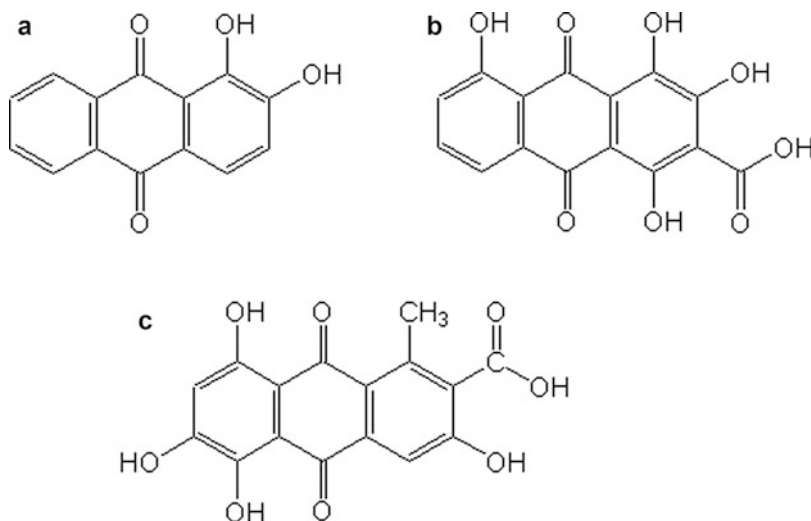
are shown in Fig. 7. Through the high number of hydroxyl groups that neighbor the quinoid system, these compounds are able to form stable metal complexes with, for example, aluminum or calcium. Thus, in traditional textile dyeing with madder extracts, mordanting with metal salts is used to improve dye fixation and fastness. The high fastness properties of such dyeings makes them valuable for textile applications, although care has to be exercised to ensure that harmful components such as lucidin (1,3-dihydroxy-2-hydroxymethyl-anthraquinone) have been removed, e.g., through an oxidation step.

Kermes, cochénille, and lac are important examples of insect-based natural colorants. In the cases of kermes and cochénille, the parasitic insects live on host plants from where the female insects are collected and dried. Lac is obtained from the secretions of the lac insect. Due to the high quality (light fastness and wash fastness) of red textile dyeings, these dyes were of high importance in traditional dyeing. The dyes still are produced in considerable amounts for cosmetic applications and as food colorants (Table 4).

Tannin-Based Dyes

Tannins are a complex group of polyphenolic compounds. Gallotannins comprise the structural unit of gallic acid esterified with sugar molecules (hydrolysable tannins), while tannins contain

Colorant, Natural,
Fig. 7 Structures of
alizarin (a), pseudopurpurin
(b), and kermesic acid (c)

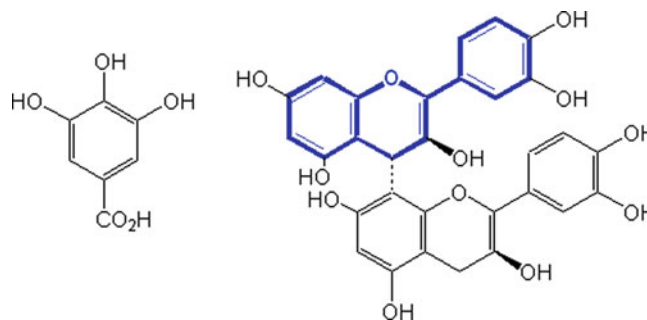


Colorant, Natural, Table 4 Insect dyes

Insect	Name	Main colorants	Color index number	Host plant
Kermes	<i>Kermes vermilio</i>	Kermesic acid	C.I. Natural Red 3	Kermes oak
Cochenille	<i>Dactylopius coccus</i>	Carminic acid	C.I. Natural Red 4	Nopal cactus
Lac insect	<i>Kerria lacca</i>	Laccaic acids	C.I. Natural Red 25	Not specific

Colorant, Natural,

Fig. 8 Structure of gallic acid and a condensed tannin ((+)-catechin-(+)-catechin, B3) flavan structure marked blue)

**Colorant, Natural, Table 5** Plant sources for gallotannins and tannin agents (condensed tannins)

Plant	Name	Part of plant	Main colorants	Content/ %	Color index number
Aleppo gall on oak tree	<i>Quercus infectoria</i>	Gallnut	Turkey tannin	50–70	
Sicilian sumac	<i>Rhus coriaria</i>	Leaves, twigs	Gallotannin	23–35	C.I. Natural Brown 6
Sticky alder tree	<i>Alnus glutinosa</i>	Bark	Gallotannin	20	
Pomegranate	<i>Punica granatum</i>	Pomegranate fruit bark	Gallotannin	28	C.I. Natural Yellow 7
Scots pine	<i>Pinus sylvestris</i>	Bark	Tannin agent	17	
Cutch/catechu	—	Bark, leaves	Tannin agent	—	C.I. Natural Brown 3
Tea	<i>Camellia sinensis</i>	Leaves	Tannins agent	25	

flavan as general structural element, by condensation complex polyphenolic structures are formed (condensed tannins) (Fig. 8).

Tannins are found in all parts of plant, e.g., bark, wood, leaves, floral parts, and gallnut. In contrast to other natural colorants, the tannin content in plant material is much higher and, depending on source, can reach levels of 50% of the plant mass (Table 5).

A gallnut is formed as response to the egg deposition by a gall wasp, e.g., on a leaves and twigs of an oak tree.

The term catechu is used for a number of plant extracts with high content in tanning agents, obtained from different plants, e.g., mangroves, conifers, cutch, and acacia.

Tannins and tannin agents form intensively colored complexes with metals such as iron and copper; thus, color development usually is coupled to use of metal mordants. An important example for such a colorant is the intensively colored iron complex with gallic acid, the dark-blue colored iron gall ink.

Tannins also are used as mordants in place of metal mordants to increase the extent of

adsorption and fastness of other natural colorants. Attractive forces in tannin mordanting then are π -bond interactions, hydrogen bonding, and Van der Waals forces.

Carotenoid Dyes

A number of yellow, orange, or red pigments are found in many plants and animals. Animals are not able to synthesize carotenoids, which they have to obtain from their food.

Carotenoids are polyisoprenoids; hydrocarbon carotenoids are carotenes, while oxygen-containing molecules are named xanthophylls. Characteristically, carotenoids comprise conjugated double bonds. Typical examples are α -carotene (carrots, *Daucus carota*; red palm oil), lutein (green leafy vegetables, broccoli, corn), bixin (annatto seeds, *Bixa orellana*, C.I. Natural Orange 4), capsanthin (paprika seed, *Capsicum annuum*), lycopene (tomato, *Lycopersicon esculentum*), and crocetin (saffron, *Crocus sativus*, C.I. Natural Yellow 6) (Fig. 9).

From annatto seeds and saffron flowers, orange dyes can be obtained by aqueous extraction. The water-soluble colorants can be used to dye wool, silk, and cellulose fibers. Both dyes also are used as food colorants.

A major group of other carotenoids exhibit poor solubility in water, and extracts are obtained using solvents of low polarity (hexane, oil). As a result, their use as food colorants is dependent on the oil content of the product as in many cases the oily phase will contain the dye.

Chlorophylls

While chlorophylls represent the most abundant pigments in nature, their use as colorants is limited due to low chemical instability and high production costs. Different natural chlorophylls have

been identified (chlorophyll *a*, *b*, *c*, *d*, and *e*). While the extraction of the green dye from plant leaves, algae, appears of high interest for the coloration of food, cosmetics, and textiles, difficulties in obtaining pure products and the rapid modification via endogenous plant enzymes to brown-green products prevent the simple, direct use of chlorophyll as a colorant.

Chlorophyllines are semisynthetic, metal-chelated chlorophyll derivatives which exhibit higher stability and are water soluble in many cases. After hydrolytic treatments and replacement of the central magnesium by copper, sodium copper chlorophyllin can be obtained. In Europe the green complex can be used as colorant for sweets, ice cream, and cheese. In the USA the use of copper chlorophyllin is more restricted, a possible use being in dentifrice.

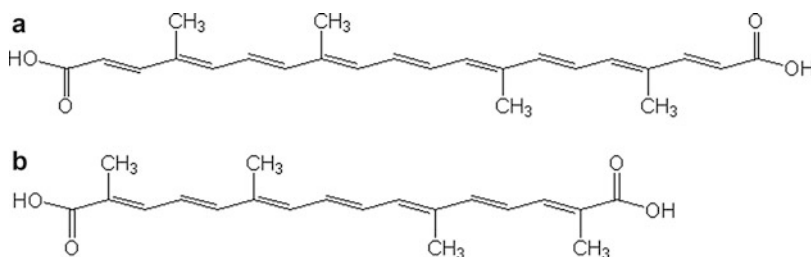
The use of copper chlorophyllin for the coloration of paper, textiles, and leather is under consideration, although the Cu content in the product has to be considered.

Applicatory Aspects

As many natural colorants are readily soluble in water, this is the main solvent used to extract the dye from a source. Due to the low content of colored material in the natural source, the amount of colorant in the extract is limited. Use of solvent extraction such as ethanol or hydrocarbons is limited to special purposes as considerable amounts of treated material, e.g., plant residues have to be deposited.

Concentration by solvent evaporation has to be considered carefully for energy reasons. Another technique to obtain concentrated dye preparations is precipitation as a dye lake by formation of insoluble complexes, e.g., by addition of calcium, iron, or aluminum salts.

Colorant, Natural,
Fig. 9 Structure of bixin
(a) and crocetin (b)



The dye lakes can be used directly as pigment dyes for paints and cosmetics.

For textile dyeing purposes, the extracted natural colorant can be used either directly without recourse to mordants. Attachment of the colorant to the substrate is then based on H-bonding, dipole interactions (Van der Waals forces).

Mordants are auxiliary chemicals used to increase dye adsorption and fixation on the substrate. The mordant can be applied in a separate bath (pre-mordanting), added directly to the dye containing bath (meta-mordanting), or used as an aftertreatment (post-mordanting).

Use of separated mordanting appears useful with regard to an efficient dyestuff consumption, however additional treatment steps and separate handling and storage of the concentrated mordanting baths make the straightforward meta-mordanting more attractive to dyehouses.

While in traditional dyeing with natural colorants, many different heavy metal salts were used (e.g., copper, tin, and chromium), environmental legislation has restricted the use of heavy metals in textile processes, and as a result, mainly iron-, aluminum-, and tannin-based mordants may be used nowadays [11].

Cross-References

- [Colorant, Textile](#)
- [Coloration, Mordants](#)
- [Coloration, Textile](#)
- [Dye, Metal Complex](#)

References

1. Nassau, K.: The fifteen causes of color: the physics and chemistry of color. *Color. Res. Appl.* **12**(7), 4–26 (1987)
2. Zollinger, H.: *Color Chemistry – Syntheses, Properties, and Applications of Organic Dyes and Pigments*. Wiley VCH, Weinheim (2003)
3. Bechtold, T., Pham, T.: *Textile Chemistry*. De Gruyter, Berlin/Boston (2019)
4. Schweppe, H.: *Handbuch der Naturfarbstoffe*. Ecomed Verlagsges, Landsberg/Lech (1993)
5. Bechtold, T., Mussak, R. (eds.): *Handbook of Natural Colorants*. Wiley, Chichester (2009)
6. Paul, R.: *Denim – Manufacture, Finishing and Applications*. Woodhead Publishing, Cambridge (2015)
7. Bechtold, T., Mahmud-Ali, A., Mussak, R.: Chapter 31. Natural dyes from food processing wastes – usage for textile dyeing. In: Waldron, K.W. (ed.) *Waste Management and Co-product in Food Processing*, pp. 502–533. Woodhead Publishing, Cambridge (2007) ISBN 1 84569 025 7
8. McClelland, R.A., McGall, G.H.: Hydration of the flavylum ion. 2. The 4'-hydroxyflavylium ion. *J. Org. Chem.* **47**, 3730–3736 (1982)
9. Derksen, G.C.H.: *Red, Redder, Madder – Analysis and Isolation of Anthraquinones from Madder Roots (Rubia tinctorum)*. Dissertation, Wageningen University, Wageningen (2001). ISBN 90-5808-462-0
10. Räisänen, R.: *Anthraquinones from the Fungus *Dermocybe sanguinea* as Textile Dyes*. Dissertation, Department of Home Economics and Craft Science, University of Helsinki, Helsinki (2002). ISBN 952-10-0537-9
11. Cardon, D.: *Natural Dyes – Sources, Tradition, Technology and Science*. Archetype Publications, London (2007)

Colorant, Nonlinear Optical

Maria Manuela Marques Raposo

Department of Chemistry, University of Minho, Braga, Portugal

Synonyms

[Dye functional](#)

Definition

Light is composed of an electromagnetic field that can interact with the elementary charges in matter, whose response can in turn influence the behavior of the other light waves. When light passes through any material, its electric field induces changes in the polarization of the material's molecules. In “linear” materials the degree of electron displacement, characterized by the linear polarizability α , is proportional to the strength of the applied electric field.

The distinguishing characteristic of nonlinear optical colorants is that their polarization response

to optical waves depends nonlinearly on the applied electric field strength. This can result in the emission of new radiation fields which are altered in phase, frequency, polarization, or amplitude relative to the incident optical radiation. Many of these effects are sensitive to specific characteristics of the local optical properties and interfaces. Multi-photon absorption can also result in electronic excitations that for a given incident light beam are more strongly localized in space than those resulting from linear absorption processes.

Nonlinear optical materials continue to attract the interest of both industrial and academic researchers due to their many versatile applications in the domain of optoelectronics and photonics.

Historical Background

Over the past decades, a variety of materials have been investigated for their nonlinear optical properties, including inorganic materials, coordination and organometallic compounds, liquid crystals, organic molecules, and polymers [1–10]. Nonlinear optics was born in 1875 when John Kerr discovered that the refractive index of glass and organic liquids could be altered by an applied electric field with the induced changes being proportional to the square of the applied field strength. Kerr measured, for the first time, the third-order nonlinear susceptibility, or what is today called the *Kerr effect* or the *electro-optical effect*. This discovery was shortly followed in 1883 by the observation of a similar but linear electric field effect in quartz; this latter process now referred to the *Pockels effect*. These nonlinear effects had limited use until the invention of the laser in 1960 by Maiman and, in the following year, the observation by Franken et al. of second harmonic generation in quartz. Following these events, the field of nonlinear optics developed explosively throughout the 1960s, highlighted by the work of Bloembergen and coworkers in exploring the full range of nonlinear optical responses of material systems and that of Buckingham and colleagues in exploring

nonlinear processes in atoms and molecules of interest to chemists [11].

Initially, optical communications technology used components almost exclusively based on inorganic NLO materials. At the time it was thought that these compounds were more suitable than organic compounds due to their greater chemical, photochemical, and physical stability and ease of processing. The two main classes of materials investigated were inorganic crystals of lithium niobate (LiNbO_3) and barium titanate (BaTiO_3) or semiconductors such as gallium arsenide (GaAs) or zinc selenide (ZnSe). Electro-optical devices that use lithium niobate have been on the market for several decades. However, these crystals have several drawbacks: high-quality single crystals are difficult to grow, are expensive, and are not easy to incorporate into electronic devices. During the 1980s it became clear that organic materials might be a better choice for use in nonlinear optical applications. A lot of organic compounds exhibit extremely high and fast nonlinearities, much better than those observed in inorganic crystals. In addition, due to the versatility of organic synthesis, their nonlinear optical properties can be modified depending on the desired application. Furthermore, organic chromophores can be incorporated into a variety of macroscopic structures such as crystals, Langmuir-Blodgett (LB) films, self-assembled films, poled polymers [1, 2, 8, 9], zeolites [12], and nanofibers [13].

At present, several organic systems (molecular as well macroscopic) with sufficiently high nonlinearities have been developed. The focus of the research in this area seems to be shifting to the optimization of a variety of other parameters, i.e., thermal and chemical stability and optical loss. Also, new approaches to design efficient nonlinear optical materials have lately emerged [7, 9].

Principles and Origin of Nonlinear Optical (NLO) Effects

In “linear” materials the degree of electron displacement, characterized by the linear

polarizability α , is proportional to the strength of the applied electric field. Nonlinear optical effects arise from the interaction of electromagnetic fields in various dielectric media to produce new fields altered in phase, frequency, amplitude, or other propagation characteristics relative to the incident optical fields with a response that depends nonlinearly on the applied electric field strength. When a beam of light propagates through a material, the electric field associated with the incident beam can provoke small displacements of the electrons within the material. This results in an induced dipole moment (μ_{ind}) which, at low electric field strengths, is linearly proportional to the amplitude of the applied electric field E , the proportionality factor being the first-order polarizability α . However, at high electric field strengths, higher-order terms become significant, and the induced dipole moment can be expanded in a Taylor series in powers of the applied electric fields (Eq. 1):

$$\mu_{\text{ind}} = \alpha E + \beta E^2 + \gamma E^3 + \dots \quad (1)$$

in which β represents the second-order polarizability or first-order hyperpolarizability and γ the third-order polarizability or second-order hyperpolarizability. This description may be expanded to the macroscopic regime of the bulk media with the first-order susceptibility and the second- and third-order nonlinear susceptibility, respectively (Eq. 2):

$$P_{\text{ind}} = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \dots \quad (2)$$

When the polarization of the incident and induced fields is taken in account, the above relations taken on a matrix form with the first and second hyperpolarizabilities (or equivalently the second- and third-order susceptibilities) being three and four rank tensors.

For practical applications, organic second-order NLO materials can be considered to be dipolar molecules organized into non-centrosymmetric lattices. Therefore, high β values are a necessary but not sufficient condition in order to obtain efficient NLO second-order materials [5, 8, 9].

Nonlinear Optical Effects and Their Practical Applications

The main NLO effects associated with linear and nonlinear susceptibilities are [1, 7, 9, 14–16]:

- First-order susceptibility $\chi^{(1)}$: *Refraction* is the change in direction of a light wave due to a change in its phase velocity. This phenomenon is the foundation of most geometric optical effects and has applications in optical fibers and lenses.
- Second-order susceptibility $\chi^{(2)}$:
 - (i) *Optical second harmonic generation (SHG)* is the nonlinear conversion of two photons of frequency ω into a single photon of frequency 2ω ($\omega + \omega \rightarrow 2\omega$) which, in the electric dipole approximation, requires a noncentrosymmetric medium. This effect was first demonstrated by Franken et al. [14] in 1961 and has applications in the frequency doubling of lasers. More recently, SHG-imaging has developed through the last decade as a progressively popular analytical technique especially for high-resolution optical microscopy for biological imaging. In fact nonlinear optical imaging has revolutionized microscopy for the life sciences due to the capacity of this technique to produce high-resolution images from deep inside biological tissues [15].
 - (ii) *Frequency mixing*: an NLO material may be used to add frequencies of two input waves to produce a single wave, whose frequency is their sum: $\omega_1 + \omega_2 \rightarrow \omega_3$ – or difference: $\omega_1 - \omega_2 \rightarrow \omega_3$. This effect has application in the frequency conversion of lasers and various nonlinear spectroscopic techniques.
 - (iii) *Parametric amplification*: an NLO material may transmit two light waves of different frequencies, with one light wave leaving the material amplified at the expense of the other. This effect allows the production coherent light at wavelengths for which no lasers are available

and is frequently used in high-resolution spectroscopy.

- (iv) *Linear electro-optical effect (Pockels effect)* is the change of refractive index of an NLO material which occurs on the application of an electric field, the extent of the change being related to the strength of the field ($\omega + 0 \rightarrow \omega$). This effect has application in waveguides and electro-optical modulators.
- Third-order susceptibility $\chi^{(3)}$:
 - (i) *Kerr effect* is a change in the refractive index of a material in response to an applied electric field. This effect has applications in optical transistors and image processing.
 - (ii) *Third harmonic generation*: light of frequency ω enters the material and leaves with frequency 3ω , ($\omega + \omega + \omega \rightarrow 3\omega$). This effect has applications in optical image processing and scanning microscopy [7].
 - (iii) *Two-photon absorption* is the induced electronic transition provoked by the near simultaneous absorption of two incident photons. This effect was first studied theoretically by Maria Goeppert-Mayer already in the 1931 and is often used in nonlinear microscopy to allow deeper penetration of the incident light, avoid photobleaching, and increase spatial resolution [16].

Nonlinear Optical Materials: Classes, Advantages, and Limitations

The search for novel chromophores for optoelectronics and nonlinear optics (NLO) is one of the main goals of modern materials science and physics. Their practical applications require not only an appropriate design but also the relevant macroscopic properties of newly established materials. The even susceptibilities χ_2 , χ_4 , etc., are only nonzero in materials lacking a center of symmetry. Therefore, the arrangements of the molecules on a macroscopic level are vitally important

to NLO activity. For example, if a crystalline material with a large value of β crystallizes in a centrosymmetric structure, the nonlinear responses of the individual molecules will cancel each other out, and there will be no resultant NLO activity. The same applies to an organic compound with an amorphous structure. If there is no order in the molecule at all, the statistical average of the NLO responses of the molecules will be zero or nearly zero, and again, the material will be not be an active NLO material. Thus to use the strong hyperpolarizabilities often found in organic molecules in a bulk phase, the constituent molecules must be somehow non-centrosymmetrically ordered. The main types of ordered assemblies that have been investigated for use in NLO materials are the following [5, 8–13].

Inorganic

These materials were the first to be studied since the observation by Franken et al. [14] of second harmonic generation in quartz. Several ionic crystals such as potassium dihydrogen phosphate ($\text{KDF}=\text{KH}_2\text{PO}_4$), lithium niobate LiNbO_3 , barium sodium niobate ($\text{Ba}_2\text{NaNb}_5\text{O}_{15}$), and β -barium borate ($\text{BBO}=\text{BaB}_2\text{O}_4$) were developed and are currently used for several optical applications such as frequency converters, electro-optical modulators, and optical switches. The main advantages of inorganic materials are their high stability, high mechanical strength, and high nonlinear optical susceptibilities. However, the growth of these crystals is time consuming (frequently requiring 1–8 weeks) and the crystals tend to be hygroscopic, and expensive, somewhat difficult to integrate into electronic devices and are often limited by the low response speeds.

Coordination and Organometallic Complexes

In the last decade, organometallic and coordination metal complexes have occupied a relevant role in the area of NLO chromophores due to an additional flexibility, when compared to organic

chromophores, owing to the presence of metal-ligand charge transfer (MLCT) transitions usually at relatively low energy and of high intensity, tunable by virtue of the nature, oxidation state, and coordination sphere of the metal center. The main classes of SHG NLO coordination and organometallic complexes include derivatives with monodentate nitrogen donor ligands (amines, stilbazoles, pyridines), chelating ligands (Schiff bases, bipyridines, phenanthrolines, terpyridines), alkenyl, vinylidene and cyclometallated ligands, macrocyclic ligands (porphyrins and phthalocyanines), metallocene derivatives, and chromophores with two metal centers (Fig. 1) [11].

Organic

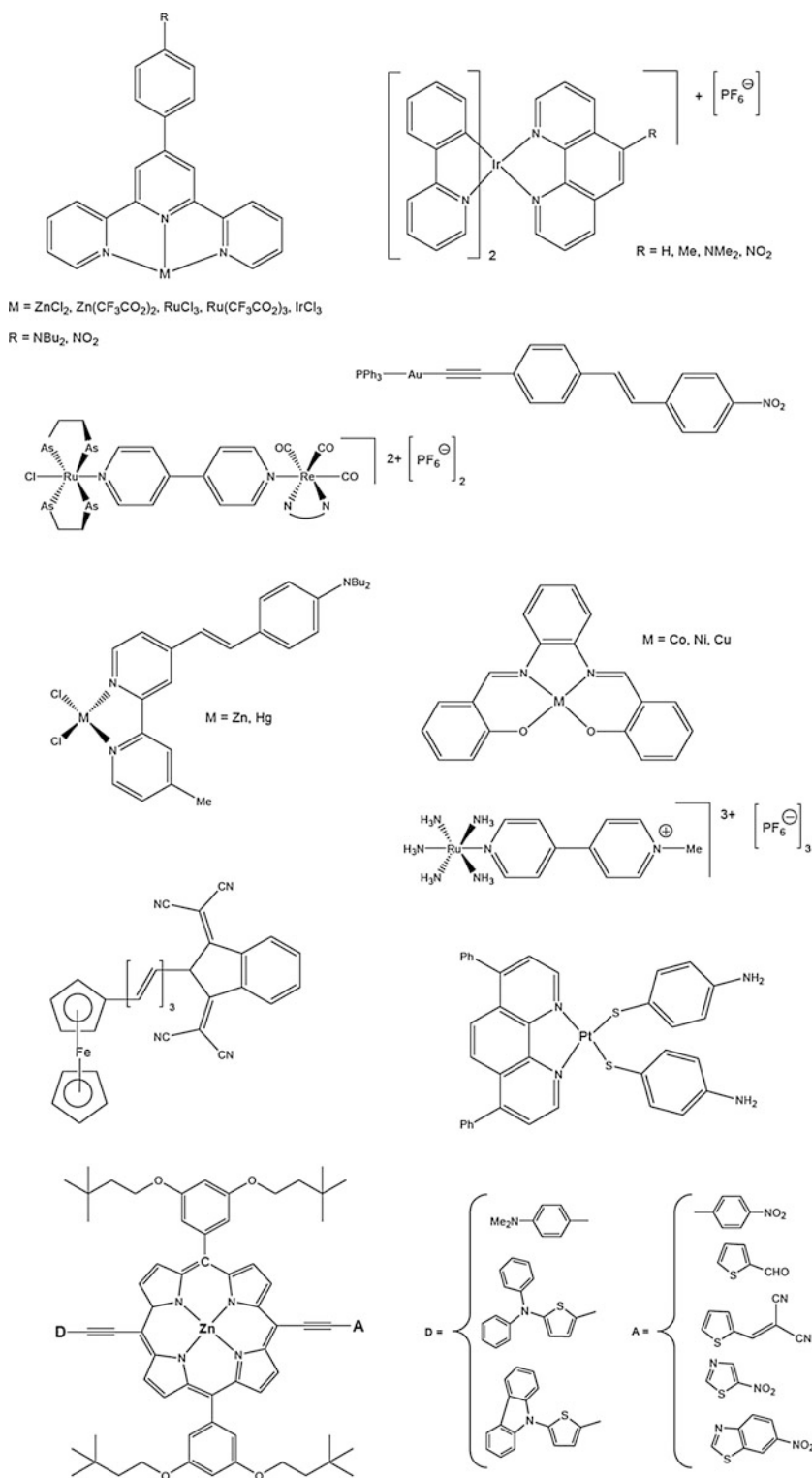
Dipolar and octopolar organic NLO chromophores (Fig. 2) have several advantages when compared to inorganic materials: (i) ultrafast response times, low dielectric constants, better tailorability and processability, as well as large NLO responses, due to the high electronic polarization of the π electrons of the molecules instead of the distortions of the atoms in the crystal lattice; (ii) easy synthesis and functionalization, allowing the optimization of their structural characteristics leading to the maximization of their NLO properties; and (iii) can be used as monocrystals and films or incorporated into liquid crystals, zeolites, or fibers. On the other hand, they have also several disadvantages such as (i) lower mechanical strength, (ii) lower photochemical stability, (iii) the propensity to acquire defects during crystal growth, and (iv) a somewhat limited working temperature range. Over the past three decades, great progress has been made in the development of new organic donor-acceptor π -conjugated systems [5, 8, 9] being driven by their chemical and thermal stability as well as the ease of synthesis and functionalization which lead to facile optical property tuning. Dipolar push-pull ($D-\pi-A$) organic chromophores are constituted by a π -bridge (aromatic or heteroaromatic) substituted with strong electron donors D (e.g., NR_2 or OR groups) and strong electron acceptor A groups

(e.g., CN, NO_2 , etc.). This $D-\pi-A$ arrangement guarantees efficient intramolecular charge transfer (ICT) between the donor and acceptor groups and generates a dipolar push-pull system featuring low-energy and intense CT absorption (Fig. 2a). The polarizability and the corresponding NLO properties, namely, the SHG of these systems, depend mainly on their chemical structure, particularly the strength of the attached donors and acceptors groups as well as the electronic nature and length of the π -conjugated bridge. However, dipolar chromophores have the tendency for unfavorable aggregation at high concentrations, and it is rare that they undergo noncentrosymmetric crystallization. One way of circumventing these disadvantages is the use of octopolar molecules.

Octopolar NLO chromophores are not so commonly investigated compared to dipolar ones. Nevertheless, their advantage is that they show the same optical transparency as their linear analogues but their second-order response is higher. Due to the fact that they possess zero overall dipole moment, they can often be oriented in the bulk phase in a manner that their molecular polarizability is additive. On the other hand, not possessing a permanent dipole moment invalidates their use in some electro-optical applications. The usual way to design SHG octopolar molecules consists in the synthesis of substituted trigonal or tetrahedral π -conjugated systems that allows an efficient charge transfer from periphery to the center of the molecule (Fig. 2b).

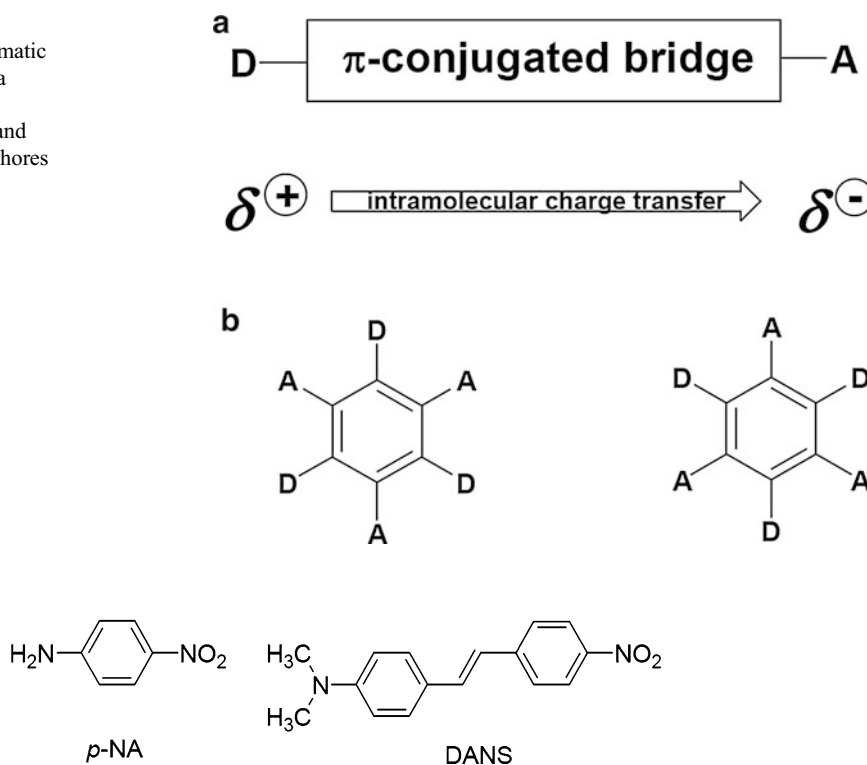
Several one-dimensional charge transfer systems with good SHG properties were developed during the 1980s. Typical examples of such molecules are *p*-nitroaniline (*p*NA) and *N,N*-dimethylaminostilbene (DANS) (Fig. 3); both are still being used as standards to evaluate SHG of other molecules.

Chromophores with other types of conjugated spacers have been also investigated (e.g., substituted benzenes, biphenyls, stilbenes, and azostilbenes). All these systems have a predominantly aromatic ground state and a corresponding charge transfer state that is quinoidal in nature. More recently, the investigation by several groups has led to the development of highly advanced SHG organic chromophores. The results of these



Colorant, Nonlinear Optical, Fig. 1 Structure of several SHG organometallic and coordination metal complexes [11]

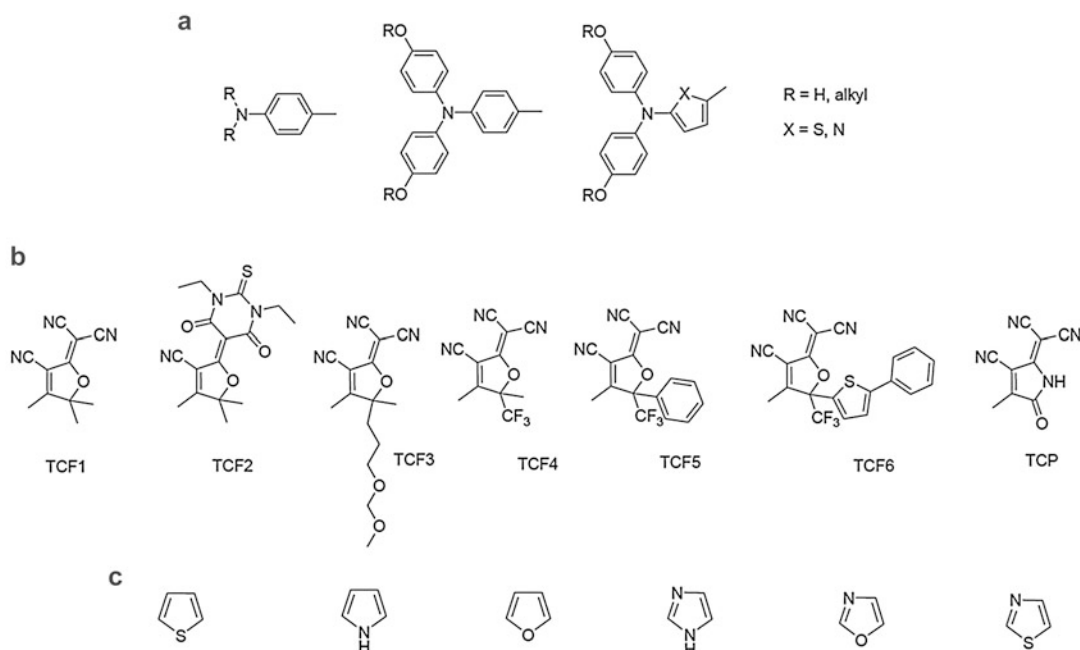
Colorant, Nonlinear Optical, Fig. 2 Schematic representations of (a) a dipolar organic D- π -A system featuring ICT and (b) octopolar chromophores



Colorant, Nonlinear Optical, Fig. 3 Structure of *p*-nitroaniline (*p*-NA) and *N,N*-dimethylaminostilbene (DANS)

studies suggest that optimal β values could be obtained through two different but complementary methodologies: the optimization of the π -bridge structure for particular donor and acceptor pairs and optimization of donor and acceptor groups for a given π -bridge. Novel highly efficient donor-acceptor pairs have been developed such as alkyl and arylamine electron donors (Fig. 4a) and tricyanofuran-based electron acceptors, for example, 2-dicyanomethylene-3-cyano-4,5,5-trimethyl-2,5-dihydrofuran (TCF1), 2-dicyanomethylene-3-cyano-4,5-dimethyl-5-trifluoromethyl-2,5-dihydrofuran (TCF4), and the tricyanopyrroline (TCP) electron acceptor (Fig. 4b). This arrangement of donor-acceptor pairs combines high chemical, thermal, and photostability [5, 8, 9]. In recent years the enhancement of β values by using ever stronger electron donor or acceptor groups has tended to reach its limit. A versatile methodology to overcome this problem was the

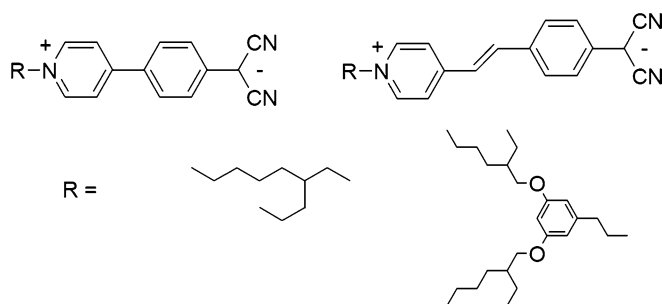
optimization of the π -bridge. Therefore, several experimental and theoretical studies have confirmed that substitution of the benzene ring of a chromophore bridge with easily delocalizable heterocycles (e.g., thiophene, pyrrole, furan, and thiazole) (Fig. 4c) results in improved molecular hyperpolarizability of push-pull systems. Due to their electronic nature and low aromaticity, they can act efficiently as π -bridges as well as auxiliary donors (electron-rich heterocycles: thiophene, pyrrole, furan) or as auxiliary acceptors (electron-deficient heterocycles: thiazole, oxazole, imidazole). In fact, the increase or decrease of the molecular nonlinear activity of these heteroaromatic systems depends not only on the electronic nature of the aromatic rings but also on the location of these heterocycles in the system. Additionally, these heterocyclic systems are also thermally and photochemically stable [8, 9, 17–20].



Colorant, Nonlinear Optical, Fig. 4 Structures of (a) aryl amines as examples of strong electron donor groups, (b) electron acceptor groups belonging to the

general TCF and TCP classes, and (c) heterocyclic compounds as π -bridges/auxiliary donor or auxiliary acceptor groups [8]

Colorant, Nonlinear Optical, Fig. 5 Structures of twisted intramolecular charge transfer molecular (TICT) chromophores [9]

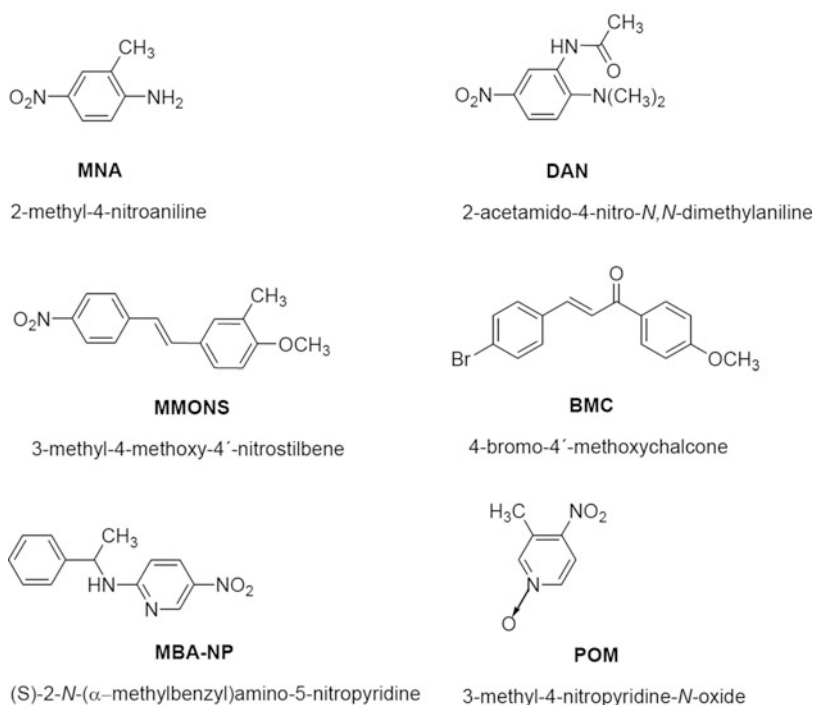


Concurrently, multidimensional charge transfer (e.g., X-shaped and higher-order symmetry) and twisted intramolecular charge transfer (TICT) chromophores (Fig. 5) have been explored as alternative approaches to improve hyperpolarizability [9].

The easiest method to impose order on the molecules of a compound is to assemble them into a crystalline matrix. Crystals have several advantages such as their high optical quality and high laser damage thresholds. On the other hand, they have several serious drawbacks; a major one is the fact that most of the promising molecules

have significant ground state dipole moments, which tend to make them crystallize centrosymmetrically. For example, 4-nitroaniline (*p*NA) packs centrosymmetrically and exhibits no appreciable SHG in crystalline form, while the analogous 2-methyl-4-nitroaniline (MNA) packs in an almost perfect head-to-tail arrangement and has a large χ_2 value. There are some examples of noncentrosymmetric small molecules such as derivatives of nitroanilines and nitropyridines or enantiomers of an optically active component. For example, MMONS

Colorant, Nonlinear Optical, Fig. 6 Structure of selected organic NLO crystals [1, 4, 5, 9]



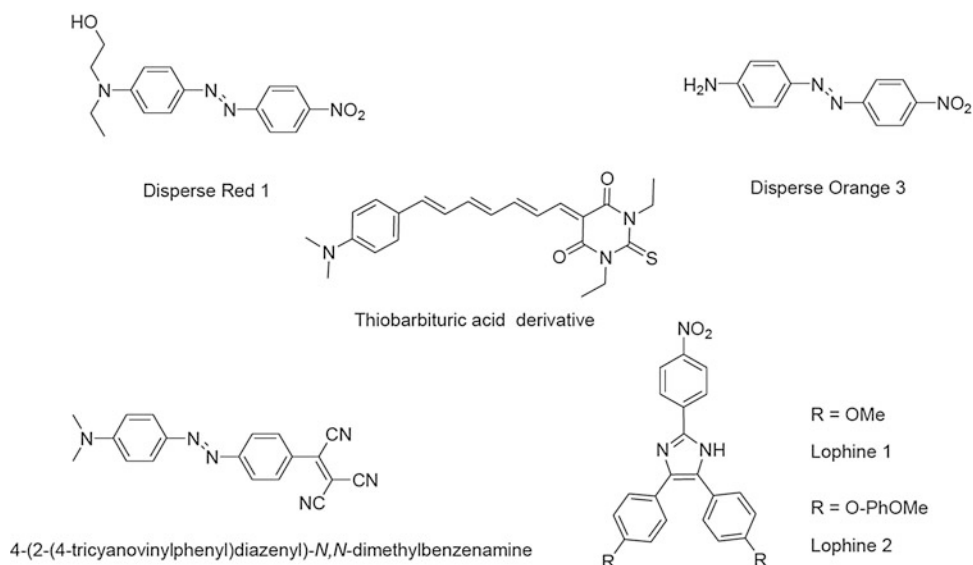
(3-methyl-4-methoxy-4'-nitrostilbene) has a very high powder SHG value (1,250 times that of urea) and POM (3-methyl-4-nitropyridine-*N*-oxide) is the only commercially available organic crystal for SHG, other than urea (Fig. 6).

Polymeric

Oriented Guest-Host Polymers

Second-order NLO applications that require crystalline materials limit the scope of molecule types that can be employed to those that crystallize in acentric space groups. To achieve good device functionality, the NLO chromophore must simultaneously possess high microscopic molecular nonlinearity, good thermal stability, good photostability, low optical absorption, and weak intermolecular electrostatic interactions in a given host matrix. Polymeric materials are attractive because they are compatible with manufacturing methods practiced in industry and can provide durability, environmental protection, and packing advantages not provided by crystalline materials.

Nonlinear optical chromophores can be incorporated into a macroscopic polymeric environment in a variety of ways. Probably the most important and most widely used is the incorporation of dipolar chromophores into a polymer host. In this approach, the active species (the guest) is dissolved in a polymeric host, which is processed to give a thin film. At this stage the molecular dipoles are randomly orientated with respect to each other. The polymer is heated above its glass transition temperature (T_g) allowing the guest molecules to become freely mobile. A strong external electric field is then applied, aligning the dipoles along one direction. With the field still applied, the polymer is cooled below its T_g again, freezing the alignment of the NLO molecules. This approach is successful in obtaining highly oriented NLO materials showing large bulk susceptibilities. The main disadvantages are (i) gradual disordering of the dipoles, (ii) limited solubility of the active species in the host polymer which limits the attainable NLO activity, and (iii) dielectrically induced breakdown of the NLO species during poling. The advantages of the poled polymer approach are the relative ease of thin film



Colorant, Nonlinear Optical, Fig. 7 Structure of selected organic chromophores used in guest-host polymers [8, 9]

making by spin coating and its compatibility with existing semiconductor technology. Figure 7 shows the structures of some examples of organic chromophores used as guests in guest-hosts polymeric systems.

Oriented Side-Chain and Main-Chain Polymers

Nonlinear optical chromophores can be also incorporated into a polymer by covalently attaching the chromophores to a polymeric backbone as part of the side chain (side-chain polymers) or by incorporation of the chromophores into the backbone of the polymer (main-chain polymers).

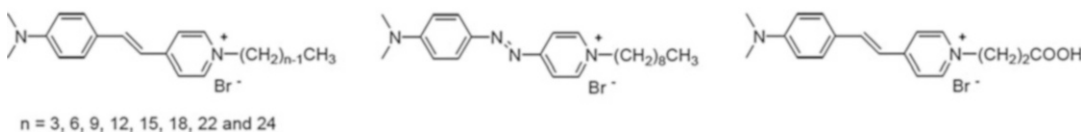
Langmuir-Blodgett Films

Another approach to organize molecules has been to incorporate organic NLO chromophores into noncentrosymmetric Langmuir-Blodgett films. The Langmuir-Blodgett film technique is used to build up ordered assemblies of molecules onto a substrate from a floating monolayer on a liquid, usually water. The molecule must be amphiphilic, that is, hydrophilic at one end and hydrophobic at the other, so that the molecules in the monolayer at

the water's surface have uniform orientation. This approach offers the advantage of much greater chromophore alignment and chromophore density. However, these films often have poor optical quality and poor temporal stability and are often very fragile.

Zeolites

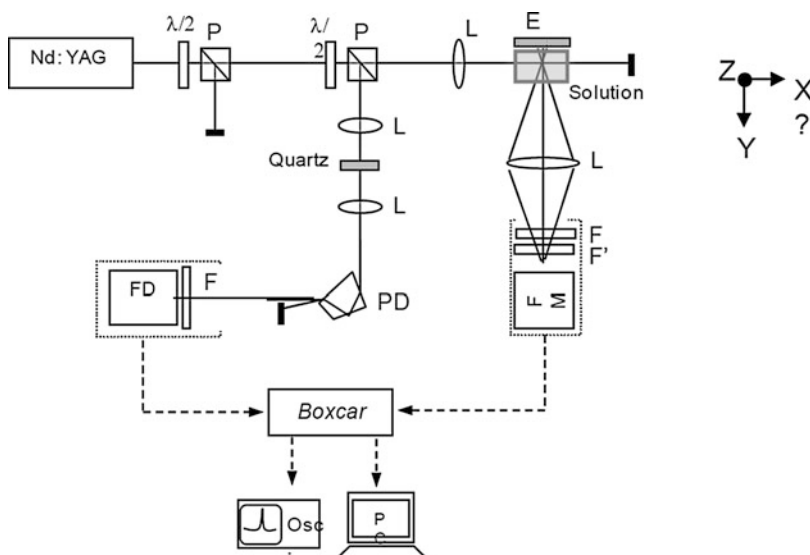
Zeolites and the related mesoporous materials have been also tested as the hosts for aligned inclusion of dipolar organic NLO dyes as a possible means to overcome the problems arising from the use of polymer matrices. The first dyes tested were *p*-nitroaniline (*p*NA), 2-methyl-4-nitroaniline (MNA), and 2-amino-4-nitropyridine which have low molecular second-order hyperpolarizability values, and the zeolites have been mostly limited to powders that bear limited practical applicability. More recently hemicyanine dyes (Fig. 8) exhibiting higher β values and a high degree of uniform orientation were introduced into transparent zeolite films with uniformly oriented channels. These second-order NLO dye-incorporating films have shown higher thermal and mechanical stability without any notable loss of activity with time. They have a



Colorant, Nonlinear Optical, Fig. 8 Structure of second harmonic generator compounds that have been incorporated into zeolite hosts [12]

Colorant, Nonlinear Optical, Fig. 9

Scheme of assembly for measurements of diffusion hyper-Rayleigh. *P* polarizer, $\lambda/2$ half-wave plate, *L* lens, *E* mirror, *PD* disperser prism, *F'* band-pass filter, *FD* photodetector, *FM* photomultiplier (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



strong potential to be practically applied in industry [12].

Nanofibers

Very recently, nanofibers of poly(L-lactic) acid (PLLA) produced by the electrospinning technique, in which donor-acceptor organic compounds such as 2-methyl-4-nitroaniline, urea, and β -glycine were imbedded, exhibit a permanent nonvanishing quadratic nonlinear susceptibility. The nonlinear optical properties displayed indicate that a noncentrosymmetric polar state was achieved and maintained a long time, allowing the use of otherwise centrosymmetric organic materials. Moreover, it was proved that the SHG efficiency of these fibers strongly depends on the diameter of the nanofibers and can be increased up to an order of magnitude by controlling the electrospinning processing parameters [13].

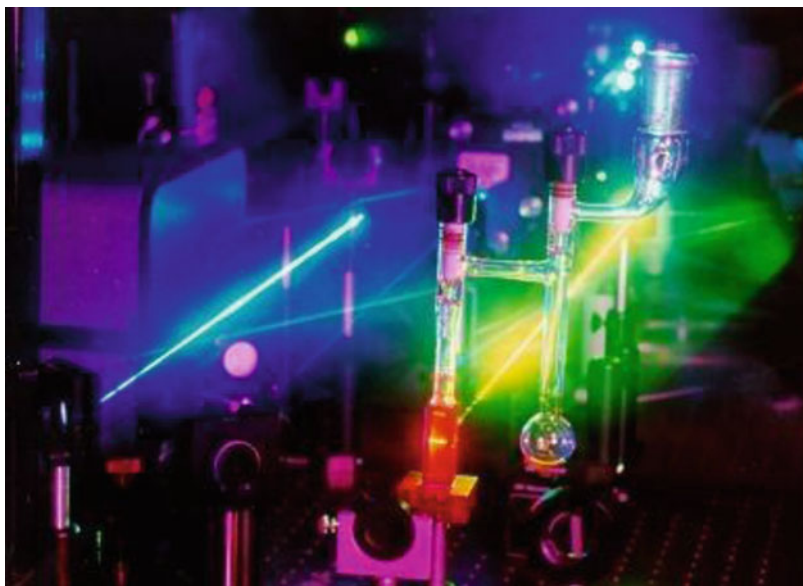
As most of the donor-acceptor organic molecules with large hyperpolarizabilities tend to crystallize in centrosymmetric structures which invalidate their use in quadratic nonlinear optical applications, the inclusion of these molecules in electrospun nanofibers may provide a means of overcoming this limitation for any donor-acceptor organic molecules with delocalized π electrons. The results of this work could have a direct impact on the design of novel nanodevices for a variety of nanophotonic applications (e.g., electro-optical transducers, pyroelectric sensors, optical frequency converters).

Experimental Methods for the Determination of Second-Order Nonlinear Effects

The second-order polarizability β and the second-order susceptibility $\chi^{(2)}$ are two parameters

Colorant, Nonlinear Optical,

Fig. 10 Experimental setup for the determination of SHG through hyper-Rayleigh scattering technique (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



indicative of a second-order response [5, 9]. The first is a molecular parameter and is usually measured in solution, whereas the latter is typically measured by second harmonic generation from the solid state. Several experimental techniques can be used in order to study these parameters in solution or in solid state: solvatochromic method, Kurtz powder technique, Maker fringes, electric field-induced second harmonic (EFISH) technique, and the hyper-Rayleigh scattering (HRS) technique (Figs. 9 and 10). In solution, the two major techniques that are presently used are EFISH and HRS.

Future and Perspectives

From the proceeding description, it is clear that the molecular design of molecules with strong nonlinear optical responses has reached a high level of sophistication. Important advances are likely to come from breakthroughs in methods that are able to make use of these strong individual molecular responses by incorporating in restricted environments that break their tendency to aggregate in centrosymmetric forms. The future of organic nonlinear optical materials is undoubtedly a bright one.

Cross-References

- [Dye](#)
- [Dye, Functional](#)
- [Dye, Metal Complex](#)

References

1. Chemla, D.S., Zyss, J.: *Nonlinear Optical Properties of Organic Molecules and Crystals*, vol. 1 and 2. Academic Press, New York (1987)
2. Prasad, P.N., Williams, D.J.: *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, pp. 132–174. Wiley, New York (1991)
3. Zyss, J.: *Molecular Nonlinear Optics: Materials, Physics and Devices*. Academic Press, Boston (1994)
4. Nalwa, H.S., Miyata, S. (eds.): *Nonlinear Optics of Organic Molecules and Polymers*. CRC Press, New York (1997)
5. Verbiest, T., Houbrechts, S., Kauranen, M., Clays, K., Persoons, A.: Second-order nonlinear optical materials: recent advances in chromophore design. *J. Mater. Chem.* **7**, 2175–2189 (1997)
6. Meyers, F., Marder, S.R., Perry, J.W.: Advanced polymeric materials – high performance polymers. In: Interrante, L.V., Hampden-Smith, M.J. (eds.) *Chemistry of Advanced Materials: An Overview*, pp. 207–269. Wiley-VCH, New York (1998)
7. He, G.S., Tan, L.-S., Zheng, Q., Prasad, P.N.: Multi-photon absorbing materials: molecular designs,

- characterizations, and applications. *Chem. Rev.* **108**, 1245–1330 (2008)
8. Cho, M.J., Choa, D.H., Sullivan, P.A., Akelaitis, A.J. P., Dalton, L.R.: Recent progress in second-order nonlinear optical polymers and dendrimers. *Prog. Polym. Sci.* **33**, 1013–1058 (2008)
 9. Dalton, L.R., Sullivan, P.A., Bale, D.H.: Electric field poled organic electro-optic materials: state of the art and future prospects. *Chem. Rev.* **110**, 25–55 (2010)
 10. New, G.H.C.: Nonlinear optics: the first 50 years. *Contemp. Phys.* **52**, 281–292 (2011)
 11. Di Bella, Dragonetti, C., Pizzotti, M., Roberto, D., Tessore, F., Ugo, R.: Coordination and organometallic complexes as second-order nonlinear optical molecular materials. *Top. Organomet. Chem.* **28**, 1–55 (2010)
 12. Kim, H.S., Cao, T., Pham, T.C.T., Yoon, K.B.: A novel class of nonlinear optical materials based on host-guest composites: zeolites as inorganic crystalline hosts. *Chem. Commun.* **48**, 4659–4673 (2012)
 13. Isakov, D.V., de Matos, G., Belsley, M.S., Almeida, B., Cerca, N.: Strong enhancement of second harmonic generation in 2-methyl-4-nitroaniline nanofibers. *Nanoscale* **4**, 4978–4982 (2012), and references cited
 14. Franken, P.A., Hill, A.E., Peters, C.W.: Weinreich: Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961)
 15. Reeve, J.E., Anderson, H.L., Clays, K.: Dyes for biological second harmonic generation imaging. *Phys. Chem. Chem. Phys.* **12**, 13848–13498 (2010)
 16. Pawlicki, M., Collins, H.A., Denning, R.G., Anderson, H.L.: Two-photon absorption and the design of two-photon dyes. *Angew. Chem. Int. Ed.* **48**, 3244–3266 (2009)
 17. Varanasi, P.R., Jen, A.K.-Y., Chandrasekhar, J., Namboothiri, I.N.N., Rathna, A.J.: The important role of heteroaromatics in the design of efficient second-order nonlinear optical molecules: theoretical investigation on push – pull heteroaromatic stilbenes. *J. Am. Chem. Soc.* **118**, 12443–12448 (1996)
 18. Breitung, E.M., Shu, C.-F., McMahon, R.J.: Thiazole and thiophene analogues of donor-acceptor stilbenes: molecular hyperpolarizabilities and structure-property relationships. *J. Am. Chem. Soc.* **122**, 1154–1160 (2000)
 19. Raposo, M.M.M., Sousa, A.M.R.C., Kirsch, G., Cardoso, P., Belsley, M., Matos Gomes, E., Fonseca, A.M. C.: Synthesis and characterization of dicyanovinyl-substituted thienylpyrroles as new NLO-chromophores. *Org. Lett.* **8**, 3681–3684 (2006), and references cited
 20. Raposo, M.M.M., Fonseca, A.M.C., Castro, M.C.R., Belsley, M., Cardoso, M.F.S., Carvalho, L.M., Coelho, P.J.: Synthesis and characterization of novel diazenes bearing pyrrole, thiophene and thiazole heterocycles as efficient photochromic and nonlinear optical (NLO) materials. *Dyes Pigments.* **91**, 62–73 (2011)

Colorant, Photochromic

Andrew Towns

Lambson Ltd., Wetherby, West Yorkshire, UK

Definition

The defining characteristic of photochromic colorants is that they change color reversibly in response to variations in the intensity of particular wavelengths of light to which they are exposed. Industry has been successfully exploiting such dyes for more than a quarter of a century, manufacturing millions of dollars' worth each year [4, 18]. The majority is consumed in the production of ophthalmic lenses that darken reversibly when exposed to strong sunshine. Photochromism continues to attract the interest of both industrial and academic researchers, who are looking to harness photochromic colorants in fields like optoelectronics and nanotechnology.

Photochromism

The widely accepted definition of photochromism is that of a reversible color change induced in a compound driven in one or both directions by the action of electromagnetic radiation [8, 9]. Photochromic systems are classified as either “P-type” or “T-type.” The former kind can be switched in each direction with different wavelengths of light. P-type systems change color when irradiated with a specific wavelength range and then remain in this state after removal of the stimulus. It is only when they are subjected to light of a different set of wavelengths that they return to their original color. In contrast, T-type behavior is exhibited if light is able to drive the change in just one direction. T-type systems will fade back to their original state, through a thermal back-reaction, when they are no longer exposed to the light source. Reversibility of response is a key aspect in both types of photochromism: light-sensitive materials that undergo changes of an irreversible nature cannot be described as photochromic.

Real-world colorants do not always match the strict definitions of the two kinds of behavior, as discussed further below. Nevertheless, most are readily categorized. Photochromic compounds of either type are available commercially. While T-type dyes are far more important industrially, there is much interest in P-type materials. Figure 1 very generally captures the behavior of both kinds.

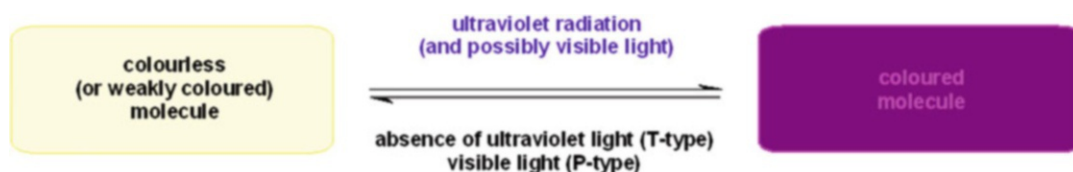
The first major application of photochromism was commercialized in the mid-1960s: glass ophthalmic lenses that relied on inorganic halide crystals. These systems have been superseded during the past three decades by organic materials in the form of plastic lenses incorporating T-type dyes [7, 17]. Such colorants are colorless – so to describe them as dyes might initially seem odd! – but they become colored when irradiated with sunlight and fade back thermally to colorless in low levels of light. Developing an economic technology with commercially acceptable durability and performance did not prove easy because photochromic colorants are less robust than conventional dyes and orders of magnitude more expensive. Light-responsive organic compounds that can be switched from one color to another are well known, but because photochromic lens manufacture remains the dominant application, the most industrially important dyes are those which exhibit T-type behavior as depicted in Fig. 1 involving colorless to colored transitions. This kind of color change, in which light causes a shift in absorption to longer wavelengths, is known as “positive photochromism.” The term “negative photochromism” does not mean non-photochromic, but instead covers the converse phenomenon of a colored dye becoming colorless

upon irradiation with light, only to return to its original colored state in the dark [1, 3].

T-Type Photochromic Colorants

There are many examples of organic compounds that change color upon irradiation with light and revert to their original state following removal of the illuminant [5, 6, 8]. The photochromism is T-type because the back reaction is driven thermally, although for commercial photochromic classes visible light may also contribute. The rate of thermal fading is often expressed as “half-life” which is the time taken for absorbance to halve once the activating light has been removed. For ophthalmic utility, a short half-life is desirable to stop vision being impaired when there is a sudden drop in light intensity [17]. However, a commercially interesting colorant for lens production must satisfy numerous other requirements:

- Weak visible absorption when unactivated so residual color is low
- Quick response to an increase in illumination
- Have a strongly absorbing activated state, because even when irradiated with light of high intensity, only a relatively small proportion of colorant will exist in this form
- Show a good compromise between depth of activated color and rate of fade to ensure both are acceptable as the properties of high intensity and short half-life tend to be mutually exclusive
- Produce satisfactory performance over 2 years by having reasonably lightfast colored and colorless forms which respond well to photostabilizers



Colorant, Photochromic, Fig. 1 General behavior of most commercial T-type and P-type photochromic colorants

- Resist the tendency to “fatigue,” whereby during activation, a proportion of the dye is undesirably and irreversibly converted to non-photochromic molecules, leading to gradual weakening of color upon repeated activation
- Color up in a manner that is not greatly affected by the temperature of its environment
- Exhibit sufficient solubility in lens media to give solutions rather than dispersions because commercial T-type classes do not exhibit useful photochromism in solid form

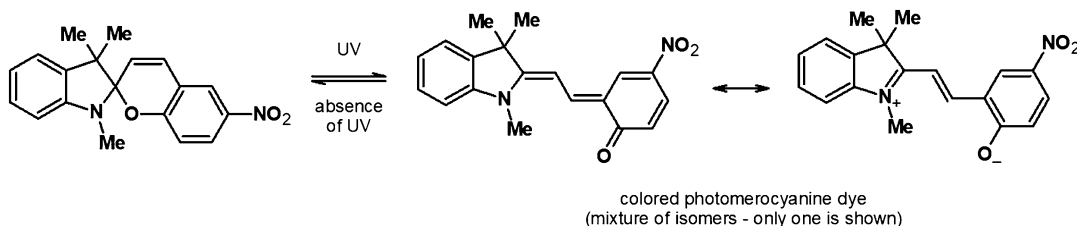
All commercial T-type dye classes undergo the same kind of molecular transformation, photoisomerization, as illustrated by an example of the well-studied spiropyran family in Fig. 2 [5, 9]. The geometries of commercial T-type colorants change substantially when switching between colorless and colored states. Consequently, the medium in which a dye is dissolved can markedly influence its photochromism. Non-polar solvents typically provide a favorable environment for photoisomerization, whereas, as discussed below, polymers can inhibit interconversion.

The wavelengths of light that effect the forward conversion shown in Fig. 2 are normally within the UVA region (315–400 nm), but blue light can also play a role for some commercial spiro-derivatives. Absorption causes rearrangement of the bonding between the atoms within a colorless or weakly colored molecule to create structures that confer intense color. The colorless form consists of a ring-closed structure, which is made up of two halves that are perpendicular to each other and joined by a spiro carbon atom. The π -systems of these moieties are small and

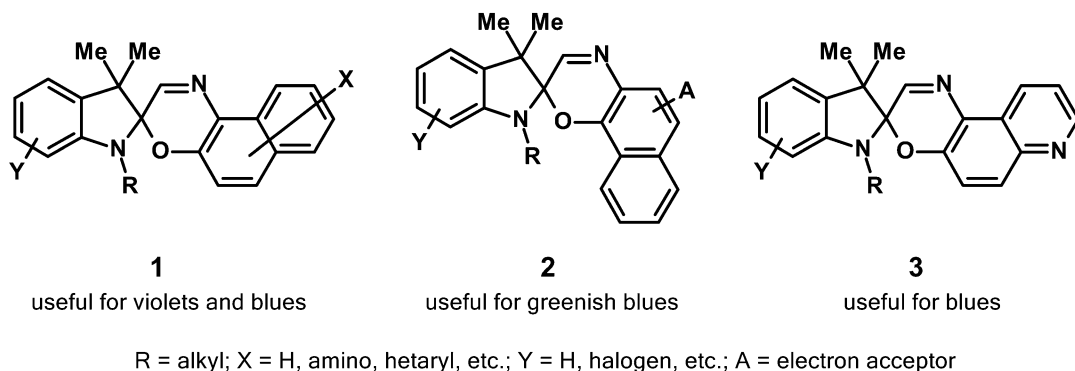
essentially isolated from each other, hence the lack of absorption in the visible region. However, absorption of energy in the form of UV light can lead to ring-opening as a result of the bond between the spiro carbon atom and its adjacent oxygen atom breaking. Molecular twisting and bending via intermediates then ensues giving planar species with extended conjugated π -systems whose absorption moves into the visible, generating color. Low light levels result in ring-closure back to the colorless form because thermal fading dominates.

Different forms of the dyes exist in a dynamic equilibrium: at a given moment, molecules are isomerizing between colored and colorless species, whose concentrations are determined by the intensity of incident light. As the flux of UV to which the dye is exposed increases, the proportion of dye that is in its colored state grows through promotion of ring-opening relative to ring-closure. Removal of the light source leads to the concentration of the colorless ring-closed form rising, which is observed as fading. When the intensity of illumination is held constant, the isomer concentrations will settle down into what is known as a “photostationary state,” where depth of color does not change. These proportions are dependent on the dye, the nature of the illumination and the medium. Because the photostationary state is a dynamic equilibrium, dye molecules will continue to swap between colorless and colored isomeric forms even after it has been attained.

Since a significant proportion of sunlight is made up of radiation in the UV, it is capable of causing pronounced photochromism. In contrast, most commercial T-type dyes do not respond well to artificial light sources, such as tungsten filament



Colorant, Photochromic, Fig. 2 Photochromism of typical example of the spiropyran class



Colorant, Photochromic, Fig. 3 Some photochromic oxazines of commercial importance

bulbs, because the proportion of their UV output is low.

Several classes of T-type compounds have been extensively investigated (e.g., anils, perimidinespirocyclohexadienones, spirodihydroindolizines, etc.) [5, 6], but few have attained commercial significance. The three families of dye that have had the greatest industrial importance will be discussed next.

Spiropyrans

This class was intensively studied during the 1950s through to the 1970s because many of its members are readily synthesized and photochrome to deep colors, typically violets and blues, that fade at useful rates [5]. However, their generally poor photostability prevented successful commercial exploitation. Nevertheless, the spiropyran unit continues to be of interest in research where light stability is not of critical importance, for example, in the fields of biochemistry and materials science [15].

Spirooxazines

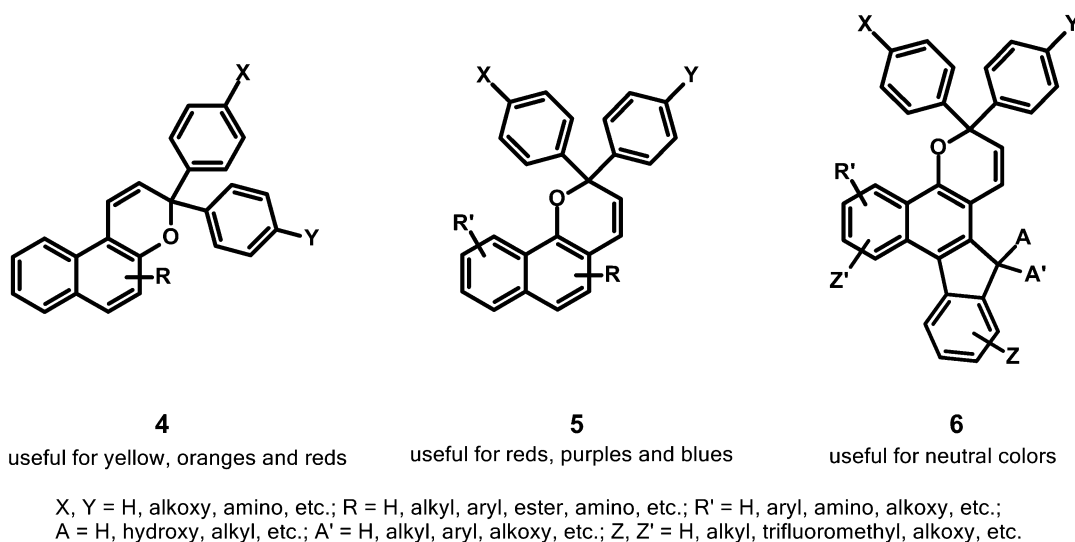
Although spirooxazines [5, 16] are similar in structure to spiropyrans, the former are much less prone to fatigue [6, 9]. As a result, they have become well established since they were employed in the production of the first commercial plastic photochromic lenses in the early 1980s [17].

Dyes derived from the indolinospirooxazine [2,1-b][1,4]oxazine **1**, indolinospirooxazine [1,2-b][1,4]

oxazine **2**, and indolinospiropyridobenzoxazine **3** classes have all have enjoyed extensive use in ophthalmic lens manufacture (see Fig. 3) [4, 18]. The simplest examples give relatively fast-fading blue photocoloration, but small adjustments to structure furnish dyes with useful intensities and half-lives. For example, placing bulky alkyl groups on the indoline nitrogen (substituent R in Fig. 3) slows down fading and increases strength. Manipulation of color is also possible with appropriate design making bluish-red through to turquoise dyes accessible commercially. Dyes **3** were successfully utilized as the blue components of a photochromic colorant mixture in the production of the first true grey photochromic lenses in the early 1990s.

Naphthopyrans

The naphthopyran class [5, 12] is the most commercially important variety of photochromic dye [4, 18]. All the major manufacturers of polymer-based photochromic lenses make use of naphthopyrans in their formulations. Their chemistry facilitates the economical production of a palette of dyes that not only spans the visible spectrum from yellows through to oranges, reds, purples, and blues, but also features more neutral colors such as olive, brown and grey. This family offers considerable scope for fine-tuning of photochromic properties because many convenient modifications to structure are possible. However, careful design is needed since substituent choice usually



Colorant, Photochromic, Fig. 4 Some photochromic naphthopyrans of commercial importance

affects both conversion kinetics and color. The stability of such dyes is generally as good as any other class, while their photochromism tends to be more independent of temperature than that of spirooxazines. The above advantages have led to naphthopyrans becoming the most commercially important type of photochromic colorant in the form of two subclasses: *3H*-naphtho[2,1-*b*]pyrans **4** and *2H*-naphtho[1,2-*b*]pyrans **5** (see Fig. 4).

Another reason for the significance of naphthopyrans is that they simplify the development of colorant recipes for lenses of the most commercially important shades, which are grey and brown [4, 18]. Incorporation of appropriate structural features into **5** produces relatively dull, neutral dyes, enabling use of fewer colorants in a mixture. Of particular industrial usefulness are “indeno-fused” dyes of general structure **6**, whose photocoloration features more than one broad absorption band in their visible spectra, furnishing murky tertiary olive, dull blue, purplish and even grey shades [18]. Given that the components of a recipe must activate, fade, and fatigue at the same rate, use of such dyes as their principal elements constitutes a great advantage to the formulator.

Applications of T-Type Dyes

While the properties of photochromic colorants have been put to purely aesthetic use in artwork, T-type dyes have also been exploited as functional materials, for example, in security printing where light-responsive marks are used as indicators of authenticity. The main outlet for photochromic dyes, lenses for spectacles, falls between these two extremes: variable transmittance is a key function in regulating light intensity reaching the eye, yet color is also an important consideration for style, comfort, and visual acuity. Although T-type colorants have been investigated for many applications, success in developing marketable products has been limited by the challenges that their usage presents. They are not as robust as conventional dyes and pigments, rendering certain methods of application unsuitable. In order to get strong, durable photocoloration, care must also be taken to provide T-type dyes with the right environment in which to operate. It must be conducive to the changes in molecular geometry associated with photochromism and not lead to rapid degradation of dye during and/or following application. These considerations will be illustrated for polymers and surface coatings, which are the two most common media for photochromic dyes.

Arresting photochromic effects can be produced by incorporation of dyes into thermoplastics. However, the chemical and physical properties of the polymer (as well as the dye) exert a large influence on the kinetics and resilience of the photochromism. For example, spirooxazines and naphthopyrans work well when used in mass coloration of polymers that have relatively low glass transition temperatures with flexible chains, such as polyethylene and polypropylene, giving striking color changes at inclusion levels of 0.3%w/w and less. However, rigid, crystalline materials restrict the necessary conformational changes for photoisomerization, severely inhibiting photochromism. Dyes can tolerate brief exposure to the high temperatures experienced in injection molding or extrusion, but certain polymers, such as polyamides, require processing at elevated temperatures that strongly degrade colorants, leading to discoloration and a lack of observable photochromism. Even when suitable polymers are employed, additives are often needed to enhance dye photostability so that the effect has a commercially acceptable lifetime. Loss of photochromism in industrial systems is related to the cumulative amount of incident UV radiation rather than the number of times that the material is switched between colorless and colored states. Consequently, UV absorbers can usefully shield dyes from excessive radiation if they do not strongly absorb the UV wavelengths which activate the dyes. Other additives include hindered amine light stabilizers that scavenge light-generated free radicals, which would otherwise attack colorants, and triplet state quenchers that inhibit photochemical reactions other than the desired one of photoisomerization.

Application to polymers need not involve monomolecular dissolution in the polymer itself. Photochromic dyes can also be used in microencapsulated solvent droplets of typically 1–10 μm in diameter. In this form, the dye solution is encased in polymer shells, producing a photochromic composite that can be dispersed like a pigment. The advantage of this material is that its photochromism is much more independent of the medium. The properties are that of the dye

dissolved in the solvent and do not tend to be influenced by the polymer in which the microcapsules are dispersed. In this way photochromic effects can be produced in substrates using the microencapsulate that would not be possible using dye alone. Disadvantages of this technique are the relatively high loadings of pigment required and the potential for microcapsules to be physically damaged during application. An alternative approach to obtaining photochromism in inhospitable polymers (as well as potentially improving the robustness and responsiveness of the effect) is to attach oligomeric fragments to dyes, for example, poly(siloxane) and poly(ether) chains [18]. These pendant oligomers are thought to provide a favorable microenvironment for the photoactive part of the colorant. All of these methodologies have been commercialized for coloration of polymers. Photochromic materials are found in products as diverse as toys, fashion accessories, fishing line, and motorcycle helmet visors [4].

T-type colorants have also been used for applications ranging from security printing to cosmetics, in the form of photochromic inks and other surface coatings. Commercial dyes work well in non-polar solvents such as toluene. When dissolved at a suitable concentration, solubility, and photochromic behavior tend to be good, enabling the formulation of solvent-based inks and varnishes. Aqueous media require an alternative approach since such dyes are not water-soluble but must be in solution to exhibit photochromism. One way is to disperse microencapsulated dyes into water-based systems using commercially available powders or aqueous slurries. Alternatively, micronized particles of an appropriate polymer into which photochromic dye has been incorporated can be employed. It is also possible to formulate inks for a variety of printing techniques, provided care is taken to ensure that dye is not damaged during formulation or application. Additives may be needed to improve light fastness. Even following optimization, photostability can be problematic. This is also true of photochromic textiles: the most effective method of applying dye to fabric is screen-printing microencapsulated colorant

because typical fibers, such as polyester, inhibit photochromism while conventional techniques, for example exhaustion dyeing, tend to damage typical commercial dyes. Photochromic detail can be added to garments through the use of polypropylene thread that has been melt-spun with dye [4].

The general lack of robustness of photochromic dyes compared to conventional colorants has prevented their use in particularly demanding applications. One example is light-responsive glass for architectural windows. Lifetimes in excess of 10 years are required but such a demand cannot be met by existing dye technology.

T-type colorants are also not suited to potentially important uses which require controlled switching between one or more states (colored and/or colorless) on demand: for these applications, P-type dyes are needed.

P-Type Photochromic Colorants

A significant amount of effort, in both academia and industry, has been invested in P-type colorants over the past three decades because of their potential as molecular switches [11]. Such compounds are converted from one state to another by irradiation with light, remaining so until switched back by other wavelengths. This behavior is of great interest in high technology sectors, but while much time and money has been spent on developing P-type applications, these endeavors have yet to bear commercial fruit.

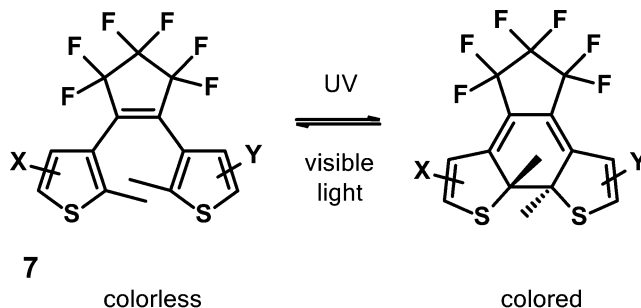
The classes that have been studied most in this connection are discussed next.

Diarylethenes

Of P-type families of colorants, this class [5, 11, 13] has arguably been subjected to the most intense scrutiny. In contrast to commercial T-type dyes, their photochromism relies upon UV light causing ring-closure rather than ring-opening (see Fig. 5).

With appropriate structural design, the colored cyclized material is essentially thermally stable (owing to a negligibly small reaction rate for thermal reversion) so that the reverse reaction does not occur in the dark. Instead, ring-opening requires the absorption of longer wavelengths of visible light than those that cause activation. Consequently, unlike pyran- and oxazine-derived colorants, diarylethenes do not fade spontaneously once the illumination is removed. A further contrast is that, whereas commercial T-type compounds must be in solution to produce useful photochromism, diarylethenes give photochromic responses in solid form (follow this [link](#) for a video). The derivatives that have attracted the greatest attention are members of the dithienylhexafluorocyclopentene subclass **7**. A few of them are available on the open market in small quantities. Activated colors can be varied widely by modification of their structure: collectively, members of this class cover most of the visible spectrum. By linking together different diarylethene skeletons, photochromic molecules have been created that can even be selectively

Colorant, Photochromic,
Fig. 5 Photochromism of
diarylethenes using the



X, Y = H, alkyl, fused (het)aryl, etc.

dithienylhexafluorocyclopentene class (7) as an example

switched between more than one different colored state.

Fulgides

Derivatives drawn from this class [2, 5, 9] have been observed to exhibit various kinds of chromic behavior of which P-type photochromism is one. Like diarylethenes, generation of color involves a ring-closure reaction upon exposure to UV radiation either in solution or as a solid. By judicious choice of structure, transitions from nearly colorless to yellow, to red, and even into the infrared region are possible. Fading is caused by absorption of certain wavelengths of visible light. An example of this class that has been commercially available is shown in Fig. 6. When activated, dye **8** changes color from pale yellow to red.

Applications of P-Type Dyes

Fulgides have been used in conventional coloration areas such as textiles and printing inks [2]. However, the greatest interest in P-type photochromic systems stems from their potential use as functional colorants within the fields of optoelectronics, data storage, nanotechnology, and medicine [10, 11, 13, 20]. In these contexts, diarylethenes have been the most intensely studied family as they offer a wide scope for the design of molecules whose optical properties can be switched in a controlled manner between persistent states. Much effort has been expended on trying to exploit this behavior, for example, in

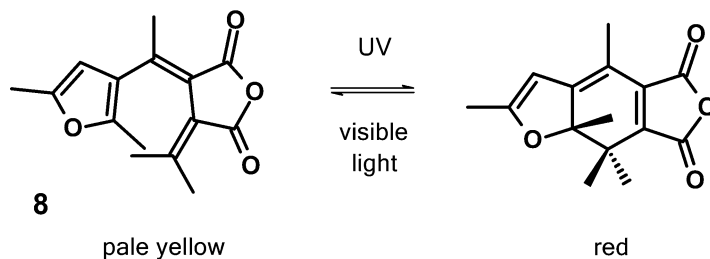
developing components for all-optical circuitry, such as switches and logic gates. Optical equivalents to electrical components, perhaps reliant on P-type colorants, are needed if the technology is to replace slower, more power-hungry conventional electronics. Molecules that can be switched optically are also being studied in the field of information technology because in theory they could furnish memory systems with much greater data storage densities than those of current commercial devices.

Another avenue of research for P-type colorants is nanotechnology because of their solid phase photochromism. For example, crystals of dihetarylethenes undergo changes in shape, as well as color, as a consequence of molecular geometry altering during photochromic transitions (follow this [link](#) for a video). Such alteration in particle dimensions could form the basis for light-driven actuators in nanomachinery.

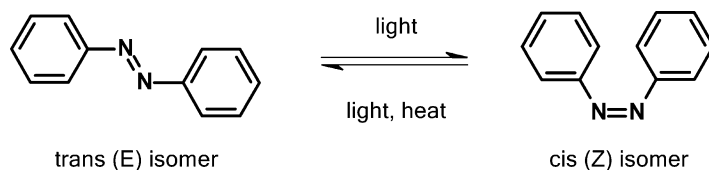
One of the oldest known photochromic systems, azobenzene, has come under increased scrutiny as a means of creating materials with light sensitive properties [21]. Azobenzene has an element of both T- and P-type character as shown in Fig. 7.

The rate of thermal back reaction can be slowed by structural modification, creating materials with strong P-type behavior where *trans* to *cis* photoisomerization occurs with one range of wavelengths and *cis* to *trans* with another set [21]. These transformations have been explored

Colorant, Photochromic,
Fig. 6 Photoisomerization of fulgides as exemplified by furylfulgide **8**



Colorant, Photochromic,
Fig. 7 Photochromism of azobenzene



for use in optical switching and data storage. The geometry change associated with photoisomerization attracts much interest as a means of creating polymeric materials that undergo reversible photomechanical deformation [14, 19]. Films which contract upon exposure to UV light and then expand when irradiated with visible light have been used to demonstrate the concept of light-driven motors (follow this [link](#) for a video).

The considerable research effort dedicated to exploiting new and old P-type photochromic systems includes Nobel Prize winning work [10]. Nevertheless, T-type photochromism remains the most commercially important phenomenon, primarily in the form of naphthopyran colorants. If one of the nascent applications of P-type dyes does however become an industrial reality, then usage of such materials would likely grow to dwarf the current volume of photochromic dyes consumed.

Cross-References

► [Dye, Functional](#)

References

1. Aiken, S., Edgar, R.J.L., Gabbutt, C.D., Heron, B.M., Hobson, P.A.: Negatively photochromic organic compounds: exploring the dark side. *Dyes Pigments*. **149**, 92–121 (2018)
2. Bamfield, P., Hutchings, M.G.: *Chromic Phenomena: Technological Applications of Colour Chemistry*, pp. 8–74. The Royal Society of Chemistry, Cambridge (2018)
3. Barachevsky, V.A.: Negative photochromism in organic systems. *Rev. J. Chem.* **7**, 334–371 (2017)
4. Corns, S.N., Partington, S.M., Towns, A.D.: Industrial organic photochromic dyes. *Color. Technol.* **125**, 249–261 (2009)
5. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds Vol. 1: Main Photochromic Families*. Plenum, New York (1999)
6. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds Vol. 2: Physicochemical Studies, Biological Applications, and Thermochromism*. Plenum, New York (1999)
7. Crano, J.C., Flood, T., Knowles, D., Kumar, A., Van Gemert, B.: Photochromic compounds: chemistry and application in ophthalmic lenses. *Pure Appl. Chem.* **68**, 1395–1398 (1996)
8. Dürr, H., Bouas-Laurent, H.: Organic photochromism. *Pure Appl. Chem.* **73**, 639–665 (2001)
9. Dürr, H., Bouas-Laurent, H.: *Photochromism: Molecules and Systems*. Elsevier, Amsterdam (2003)
10. Feringa, B.L.: The art of building small: from molecular switches to motors (Nobel lecture). *Angew. Chem. Int. Ed.* **56**, 11060–11078 (2017)
11. Feringa, B.L., Browne, B.R. (eds.): *Molecular Switches*. Wiley-VCH, Weinheim (2011)
12. Hepworth, J.D., Heron, B.M.: Photochromic naphthopyrans. In: Kim, S.-H. (ed.) *Functional Dyes*, pp. 85–135. Elsevier, Amsterdam (2006)
13. Irie, M., Fukaminato, T., Matsuda, K., Kobatake, S.: Photochromism of diarylethene molecules and crystals: memories, switches, and actuators. *Chem. Rev.* **114**, 12174–12277 (2014)
14. Jerca, F.A., Jerca, V.V., Hoogenboom, R.: Photoresponsive polymers on the move. *Chem.* **3**, 533–536 (2017)
15. Kortekaas, L., Browne, W.R.: The evolution of spiropyran: fundamentals and progress of an extraordinarily versatile photochrome. *Chem. Soc. Rev.* **48**, 3406–3424 (2019)
16. Lokshin, V., Samat, A., Metelitsa, A.V.: Spirooxazines synthesis, structure, spectral and photochromic properties. *Russ. Chem. Rev.* **71**, 893–916 (2002)
17. McArdle, C.B. (ed.): *Applied Photochromic Polymer Systems*. Blackie, Glasgow (1992)
18. Towns, A.D.: Industrial photochromism. In: Bergamini, G., Silvi, S. (eds.) *Applied Photochemistry*, pp. 227–279. Springer, Dordrecht (2016)
19. Wang, X.: Photoresponsive Azo polymer films. In: Mark, H.F. (ed.) *Encyclopedia of Polymer Science and Technology*. Wiley (2017). <https://doi.org/10.1002/0471440264.pst651>
20. Wang, L., Li, Q.: Photochromism into nanosystems: towards lighting up the future nanoworld. *Chem. Soc. Rev.* **47**, 1044–1097 (2018)
21. Zhao, Y., Ikeda, T. (eds.): *Smart Light-Responsive Materials: Azobenzene-Containing Polymers and Liquid Crystals*. Wiley, Hoboken (2009)

Colorant, Textile

Andrew Towns
Lambson Ltd, Wetherby, West Yorkshire, UK

Definition

Textile colorants impart color to a textile material, usually with a high degree of permanency, as a result of their chemical binding or physical entrapment within or around the textile material. The

textile material may be in one of several forms such as fiber, yarn, fabric, garment, etc. Textile colorants are supplied in both solid and liquid forms, for example, as powders, granules, solutions, or dispersions. In certain instances, precursors are applied to textile materials to generate the colorant in situ within the textile.

Textile Dyes and Pigments

Both dyes and pigments are used in the coloration of textiles [1, 2]. The former substances are present in solution at some point during their application, whereas the latter colorants remain insoluble within any vehicle in which they are applied as well as within the textile material itself [3]. Pigments must therefore either be incorporated into textile fibers during their construction (e.g. mass coloration of a polymer followed by its melt extrusion into fibers) or be printed onto a fabric as part of a formulation that contains a binder so that the colorant is physically held to the surface of the textile in a coating. Textile dyes are usually applied from solution although certain types may be present initially as a dispersion or applied from the vapor phase. The mechanism by which dyes remain within a textile depends on the particular colorant type. Retention may rely on intermolecular forces operating between dye and fiber following adsorption onto and/or dissolution within the polymer, formation of covalent bonds between the dye and the fiber, or entrapment of colorant particles within the textile by deposition of an insoluble form of the dye.

Natural Textile Colorants

Prior to the synthesis of picric acid in the eighteenth century as a yellow dye for silk, all textile colorants were obtained directly from natural sources, such as plants, insects, and shellfish [1, 4]. These natural colorants began to be superseded by synthetic dyes and pigments in Europe during the second half of the nineteenth century. These early artificial products enabled dyers to create a wider gamut of shades and more brilliant hues

than had hitherto been possible. While there has been a revival of interest recently in natural textile colorants driven by perceptions of renewable sourcing and low environmental impact, they are less suited to high-volume industrial use for a variety of reasons. For example, they offer only a limited gamut and generally display moderate levels of fastness at best. In addition, natural dyes often require the use of a fixative, known as a *mordant*, to achieve satisfactory permanency; traditional metallic mordants are environmentally unfriendly. Textile dyes that were originally obtained from natural sources, such as Indigo, are more efficiently obtained by chemical synthesis [1, 5]. Such substantial areas of land are needed to cultivate crops from which to harvest plant-based colorants that it would not be possible to meet the current global demand for textile dyes. Consequently, while textile coloration with natural dyes is practiced commercially, it remains a low-volume activity at the high value end of the market.

Synthetic Textile Colorants

The vast majority of textile colorants are synthesized chemically on an industrial scale [5, 6]. Over a million tons are produced globally each year. Tens of thousands of colorants have been marketed since the first commercially successful synthetic textile dye, *Mauveine*, was manufactured in the late 1850s [1, 7]. The financial rewards from this particular colorant and its contemporaries spurred on much endeavor into making synthetic dyes. The genesis of the modern chemical industry, which dwarfs the colorant sector and now manufactures products as diverse as plastics, pharmaceuticals, and drugs, lies in the attempts of mid-nineteenth century chemists to prepare textile dyes and other colorants. These efforts initially involved a trial-and-error approach since little was known about molecular structure and dyes often reached the market as mixtures of different compounds with their inventors having little idea of their composition. However, a more systematic approach to textile colorant research, principally led by the German

dye industry, resulted in a better understanding of both chemistry and color–structure relationships. At the beginning of the twentieth century, Germany manufactured 85% of the world’s synthetic dyes with 10% being produced in Britain, France and Switzerland. Over a century later, manufacture is concentrated in Asia, particularly China and India, although some of the biggest suppliers remain headquartered in Europe. This shift was originally driven by the lower cost bases of eastern producers relative to those of existing manufacturing sites in Europe and North America. Contributors to the disparity included reduced capital costs, cheaper labor, and less stringent environmental constraints than were prevalent in the West at the time. The global market for textile colorants is worth several billion dollars [3]. The textile industry is the largest consumer of dyes and organic pigments. Some of the biggest-selling textile dyes are truly commodities, being produced in volumes of over one thousand tons each year for supply at just a few dollars per kilogram or even less.

Nomenclature and Composition of Commercial Textile Colorants

Traditionally, the trade names which manufacturers give to colorants are typically comprised of three parts [5]. The first element is a brand name, often incorporating elements of the producer’s name, the type of colorant, and/or its intended use. The next part indicates general color, occasionally with modifiers to highlight shade or application characteristics to which the manufacturer wants to draw attention. The last part of the name is made up of a code of letters and numbers for further differentiation of the colorant from others – these may be related to color, application properties, as well as strength. For example, the now-defunct manufacturer, *Holliday Dyes and Chemicals*, marketed the dye *Polycron Yellow C-5G 200%*. “Polycron” was the name shared by the company’s disperse dyes developed solely for the coloration of polyester; “C” denoted the disperse dye application category into which this colorant fell; “5G” indicated that it had a

relatively greenish hue; the suffix “200%” signified that its tinctorial strength was double that of the market norm (“100%”), i.e. twice the amount of active dye was present per unit mass of colorant. Many other companies employ this system of nomenclature for their colorants. However, reliance on commercial names alone will not necessarily enable informed colorant selection (i.e. which colorant to use for a particular substrate and application technique), nor show which textile colorants in the market are equivalent. Fortunately, many manufacturers link their products to the *Colour Index* (CI) generic naming system which assists users in making sense of the vast array of textile colorants that are commercially available. (Note the British English spelling of color in the title as the nomenclature scheme was originated in the UK.) For instance, Polycron Yellow C-5G 200% has a CI Generic Name assigned to it, CI Disperse Yellow 119, thereby allowing users to identify equivalent products. While colorants with the same CI Generic Name ought to contain the same main colored compound, they may not be exactly equivalent. Variations in impurity content, whether inadvertent or through deliberate inclusion of shading components, may affect color or other application properties. Differences in physical form can affect performance. Members of several types of dye class typically include substantial quantities of additives, present as processing aids, such as dispersing agents or buffers to stabilize pH. Another source of variation is the presence of a diluent, for example salt or dextrin, which is added to standardize certain dye types to a desired tinctorial strength. A further potential complication is that some textile colorants do not have a CI Generic Name, either because they are mixtures of colored components as is often the case with navy and black dyes, or simply because one has not been disclosed or assigned.

Textile Colorant Structure

The constitution of commercial dyes and pigments used for textile coloration can be described as lean: each part of a colorant molecule has one or

more purposes as will be illustrated in some of the following sections. These functions may be related to adjustment of color (e.g., hue, intensity, brightness), physical character (e.g., solubility, crystal structure, volatility), dyeing behavior (e.g., substantivity, leveling), fastness (e.g., light, heat, moisture) and so on, although there will be occasions when substituents are present merely for convenience or cost [7]. Often, dye design involves an element of compromise as certain properties will have a degree of mutual exclusivity, e.g. rapid dyeing and good leveling behavior at the expense of high wash fastness. The most general structural classification centers on the key molecular features of a colorant that gives rise to its color. Azo derivatives are the largest such class of textile colorants, although there are many others of significance.

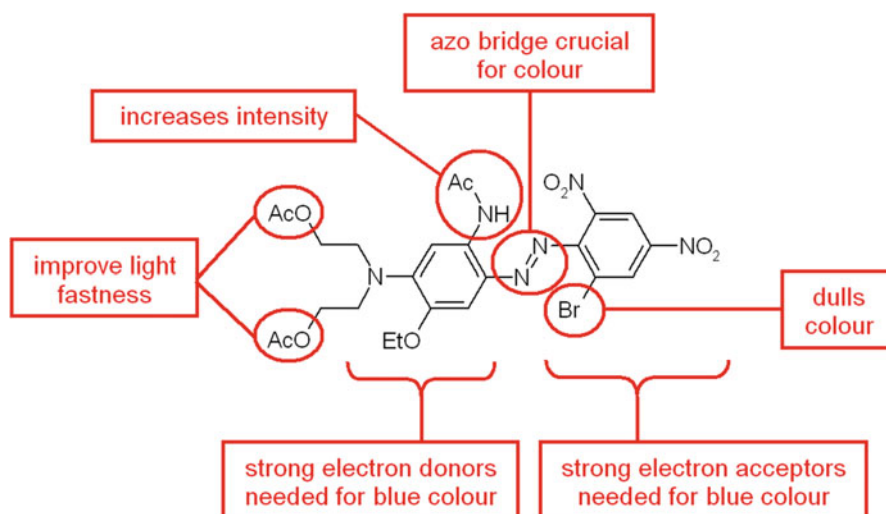
Azo Textile Colorants

This class is defined by the presence within the colorant molecule of an arylazo group of general structure ($\text{Ar}-\text{N}=\text{N}-\text{X}$) where X is most commonly another aryl ring [5–8]. In many industrial colorants, the azo function exists exclusively in a more energetically stable hydrazone form ($\text{Ar}-\text{NH}-\text{N}=\text{X}$) instead [1]. Coverage of the whole visible spectrum is possible using

commercial dyes containing just a single azo or hydrazone bridge, although many important textile dyes contain two or more such groups [1, 5, 7, 8]. Azo colorants make up around half the number of textile dyes and pigments available. This dominance lies in their economy, robustness, and versatility. Not only are they relatively inexpensive to manufacture, but generally azo derivatives also have good tinctorial strength and so tend to be economical compared to other colorant types. The class has also offered manufacturers much scope to adjust structure, enabling them to readily modify properties including shade, solubility, dyeing behavior, as well as fastness [7]. Figure 1 illustrates how structure can be divided into fragments with different purposes using the highest volume navy blue dye for polyester (CI Disperse Blue 79) as an example.

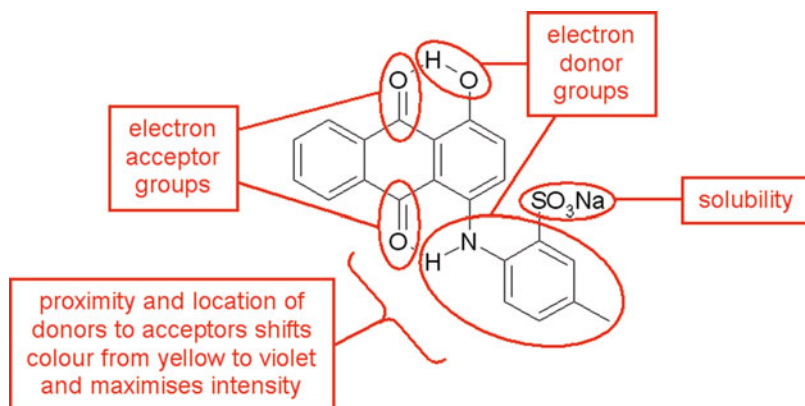
Carbonyl Textile Colorants

The color and application properties of this group of dyes and pigments are dependent on carbonyl functions ($>\text{C}=\text{O}$) [1, 5, 7, 8]. A major subclass is based on 9,10-anthraquinone, because it is a source of red to blue, as well as green, dyes of high brightness and fastness. However, economic considerations restrict their use - they are generally more expensive to manufacture than azo



Colorant, Textile, Fig. 1 Structural features of CI Disperse Blue 79

Colorant, Textile,
Fig. 2 Structural features
 of CI Acid Violet 43



derivatives, while their relatively low tinctorial strength compared to other classes means that more dye must be used to achieve a particular depth. An example of a commercial anthraquinone dye is shown in Fig. 2.

This group also encompasses one of the oldest types of textile dyes, *indigoid*, and one of the newest, *benzodifuranone*. The latter is shown in Fig. 3 and is a commercially useful source of intense bright red dyes [5].

Other Textile Colorant Structural Types

There are numerous other structural classes of textile dyes [1, 5, 7, 8] and pigments [6, 7], each occupying its own application niches. A few examples are presented in Fig. 3. The *triarylmethine* class dates all the way back to the 1850s, although the importance of such purple, blue and green dyes has diminished with the development of alternative chemical types of superior fastness. The same is true of yellow to red fluorescent *xanthene* dyes as textile colorants. However, the *phthalocyanine* class remains as much prized as ever since its commercialization in the first half of the twentieth century. This colorant type furnishes robust and tinctorially strong bright blue to green dyes and pigments with diverse uses in the textile arena, especially when they are in the form of copper complexes [5, 6]. Other classes occupy more specific sectors in terms of color and/or application. For example, *triphenyldioxazines* are exploited as intense bright blue dyes for natural fibers, while *nitrodiophenylamines* are cost-effective yellow colorants of

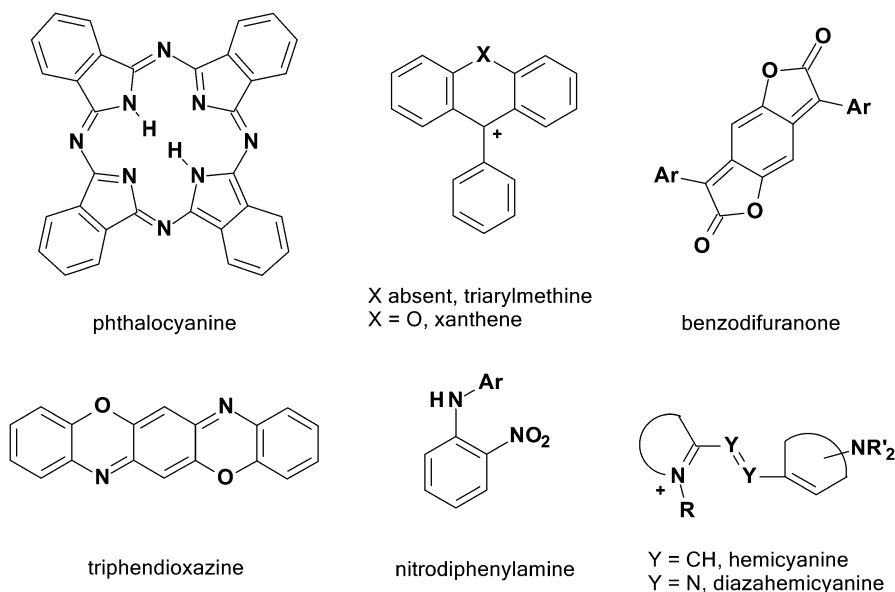
good photostability for synthetics [5]. Polymethine dyes are a broad class of chromophore of which certain subclasses are employed in textile coloration: two such types, *hemicyanines* and *diazahemicyanines*, are shown in Fig. 3. Both families are useful for intense bright reds, while the latter also supplies blue colorants [5, 8].

Textile Colorant Application Classes

Although knowledge of the molecular structure of a textile colorant is essential for a manufacturer, the user is concerned more with its method of application and performance [3, 9]. Modes of dye and pigment use include mass coloration, exhaustion dyeing, thermofixation, and printing by screen, inkjet or sublimation – each calls for a different mix of behaviors, while commercial demands will dictate economic and quality criteria [10]. A useful way of grouping textile colorants into sets with very broadly similar characteristics in terms of color, usage, fastness and economy is by CI application class. The most important application classes of textile colorant are listed in Table 1 and described in more detail below [1, 2, 5, 7–10].

Acid Dyes

This type of anionic colorant is commonly used to dye and print natural (wool, silk) and synthetic (nylon) polyamides [2, 5, 7–10]. The class is named after the acidic (pH 2–6) dyebaths used for dye application. Lowering pH increases the concentration of cationic ammonium groups



Colorant, Textile, Fig. 3 Some important textile colorant structural types

Colorant, Textile, Table 1 Major textile colorant application classes

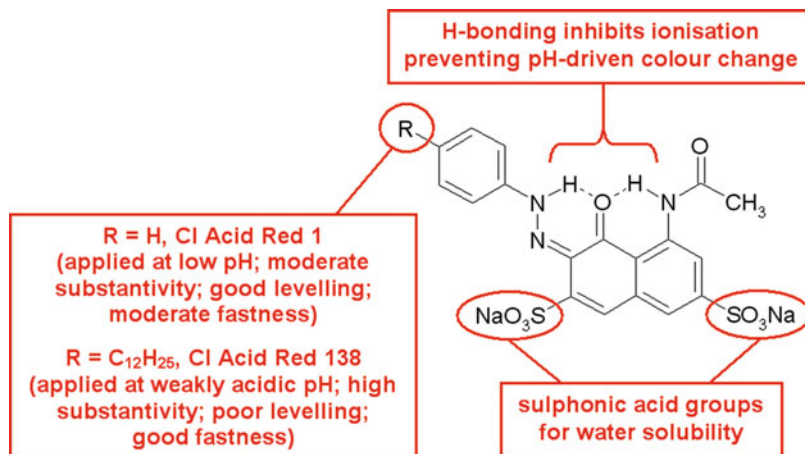
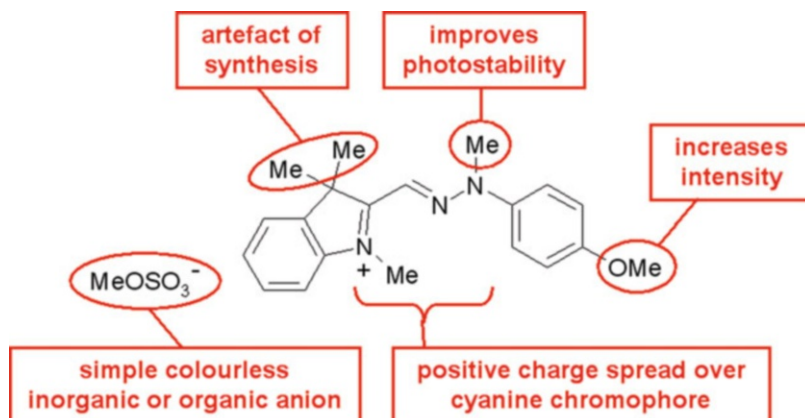
Class	Principal textile substrate
Acid	Wool, silk, nylon, modified polyacrylonitrile
Azoic	Cotton and other cellulose, acetates
Basic	Polyacrylonitrile, modified nylon and polyester
Direct	Cotton and other cellulose, polyamide
Disperse	Polyester, acetates, nylon, polyacrylonitrile
Pigment	Cotton, polyester
Reactive	Cotton, wool, silk, nylon
Sulfur	Cotton and other cellulose
Vat	Cotton and other cellulose, polyamide

within these substrates, enhancing their attraction for anionic dyes. While anthraquinone and triarylmethine derivatives are significant for violet, blue and green colorants, azo compounds are by far the most important structural class. A substantial proportion of acid dyes are metal complexes that comprise one or two dye ligands; of particular importance are 1:2 chromium:dye ligand complexes as these furnish dull, deep shades of good wet fastness and high photostability. Dyeing performance and fastness

properties are readily modified by adjusting dye hydrophobicity (see Fig. 4).

Azoic Colorants

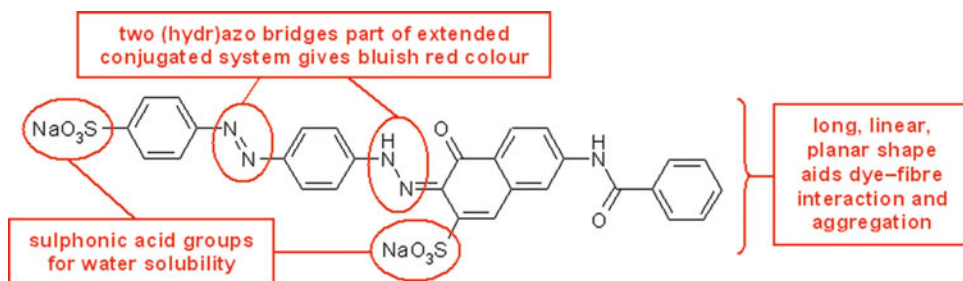
These colorants are employed predominantly on cellulosic fibers [8, 10]. They are water-insoluble azo compounds that are synthesized within the textile substrate during the dyeing process from soluble precursors. The colorants generated are physically trapped as solid particles within fiber pores, so dyeings display good fastness to laundering and light. There are numerous variations in technique and materials, but the usual method is to apply a naphthol coupling component followed by a diazo component, which reacts with the coupler to produce a non-ionic colorant. The diazo component is an aromatic amine, which is diazotized as part of the application process, or supplied as stabilized pre-diazotized materials, for example, aryldiazonium tetrafluoroborate salts (i.e., $\text{ArN}_2^+\text{BF}_4^-$). Azoic textile colorants have largely been sidelined industrially owing to the greater convenience and economy of other systems, but are still of commercial interest in the red shade area.

Colorant, Textile,**Fig. 4** Structural features of CI Acid Red 1 and CI Acid Red 138**Colorant, Textile,****Fig. 5** Structural features of CI Basic Yellow 28**Basic Dyes**

Members of this class are used primarily for coloration of polyacrylonitrile, either by exhaustion dyeing or during fiber production [2, 5, 7]. They are often referred to as cationic dyes because their molecular structures feature a positive charge as shown for example in Fig. 5. This charge may be pendant (i.e., localized but isolated from the chromophore) or delocalized, forming part of the chromophore itself [8]. In either case, the dye binds to the substrate by electrostatic interaction. Triarylmethine colorants are significant in the blue and green sectors, while azo and cyanine-type dyes dominate the yellow to violet areas [7]. All can furnish bright intense shades of high wet and light fastness.

Direct Dyes

This class is so named because its members can be applied to cotton and other cellulosic fibers without the need for mordants [2, 5, 7]. The crucial molecular features of direct dyes are (a) long, linear planar geometries and (b) multiple substitution with negatively-charged sulfonic acid groups as depicted in Fig. 6. Sodium chloride or sulfate is added to the dye bath to enhance dye adsorption. Dye geometry enables close approach to the polymer chains of the substrate. Intermolecular forces that operate at short distances can therefore become significant, aided by the large surface area of the molecule. While diffusion rates of dye into fiber tend to be low because of their large molecular size and propensity to aggregate, diffusion rates out of



Colorant, Textile, Fig. 6 Structural features of CI Direct Red 81

the dyed cellulose during washing are also small so that wash fastness is moderate. The class is dominated by azo derivatives: disazo dyes for yellows to blues and polyazo colorants for blue, green and neutral shades. Direct dyes are used for their economy where wash fastness is not critical [8, 10].

Disperse Dyes

These non-ionic colorants are hydrophobic like the synthetic textile substrates to which they are applied [9, 10]. Originally developed for cellulose acetate fibers, their principal use is the coloration of polyester: the importance of this textile material means that disperse dyes have become one of the two most important types of textile colorant [5, 8]. They have only sparing water solubility and so are applied as fine dispersions in water (apart from when they are used in transfer-printing ink films). Dye particle sizes are typically distributed in the range of around 0.1–1 μm diameter for supply either in solid or liquid dispersion form. These forms are prepared by milling dyes as a suspension in an aqueous solution of dispersing agent. The latter is usually an anionic polymeric material, such as a lignosulfonate or an arylsulfonic acid condensate, that inhibits re-aggregation and maintains dispersion stability. To obtain the micronized dye in solid form, the aqueous dispersion is spray-dried after milling in order to remove the water. During exhaustion dyeing, a small proportion of colorant is in aqueous solution: it is from this phase that dye is adsorbed onto the textile and diffuses within it. Colorant lost from the aqueous phase is replenished by dissolution of dye remaining in suspension. The micron sizing of dye particles is essential in ensuring that

colorant dissolves at a fast-enough rate and to a sufficient degree. Disperse dyes must have relatively compact structures to enable them to diffuse satisfactorily within hydrophobic textiles, so commercial ranges consist mainly of monoazo and anthraquinone derivatives: the latter are important in the bright red and blue sectors, but the former dominate the rest of the gamut. Several other chemical dye classes are employed but these tend to occupy niches, such as yellow nitrodiphenylamines. Brown and black disperse dyes are usually formulated using mixtures of azo dyes because of the technical and economic difficulties in creating single component colorants of appropriately small molecular size with the desired hue.

Pigments

While used primarily for the coloration of media other than textiles, pigments can be printed onto fabrics. Examples of organic pigments used as textile colorants in this way include yellow acetanilide and red naphthol monoazo derivatives as well as blue copper phthalocyanines [6]. The insoluble colorant is incorporated into a paste, which is applied to the textile fabric by a printing process (e.g. screen, roller, etc.). After curing of the paste, normally achieved by drying, the pigment is physically bound to the fabric. Pigments are also employed in the mass coloration of many types of textile fibers in which the colorant is dispersed in a solution or melt of the polymer prior to fiber formation so that the pigment particles become trapped within the fibers as they are produced. The pigment types mentioned above as well as numerous other chemical classes are employed in this capacity.

Reactive Dyes

These colorants are the only ones that form covalent bonds with the textile to which they are applied [2, 5, 7–10]. While they can be applied to wool, nylon, and silk, reactive dyes for cotton have achieved the greatest commercial success, becoming one of the two most important application classes. After exhaustion or printing onto the cellulosic substrate, the dye reacts with ionized hydroxy groups, furnishing excellent wash fastness. Reactive colorants resemble direct dyes with one or more attached fiber-reactive groups. This geometry favors adsorption and reaction with cellulose, although the reliance on covalent bonding allows greater flexibility in choice of chromogen. The vast majority of reactive dyes are azo derivatives apart from in the bright blue and green shade areas where anthraquinone-, phthalocyanine-, and triphendioxazine-based colorants among others are important [5, 7]. Commercially important reactive groups bond to the substrate either by addition, e.g., vinylsulfone anchors, or by substitution, such as through halogenotriazine functions like that illustrated in Fig. 7.

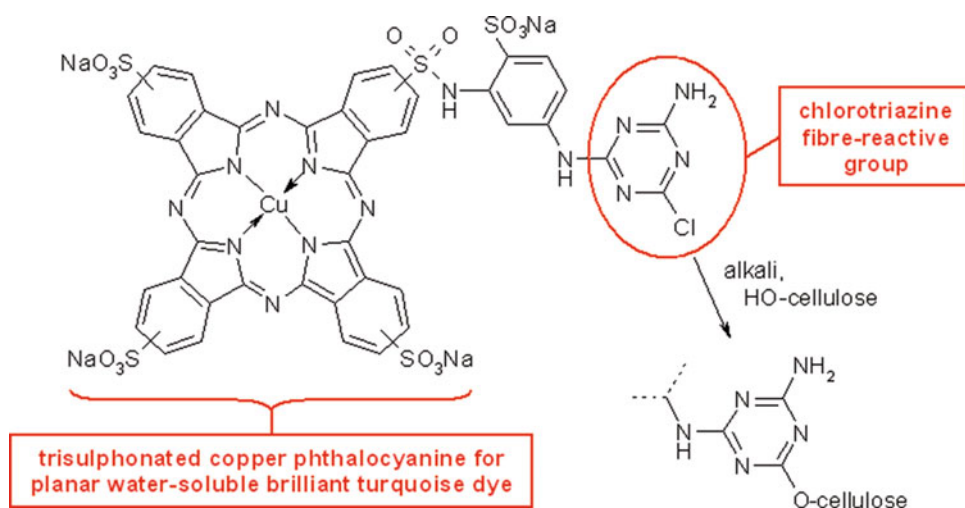
Sulfur Dyes

These colorants are prepared by heating aromatic compounds with sulfur or a sulfur compound to

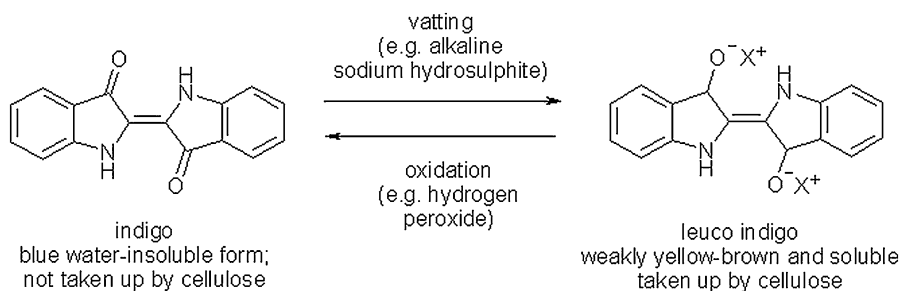
produce compounds of large molecular size that contain disulfide ($-S-S-$) linkages between the aromatic components [1, 5, 7]. Dyes of this kind are chemically complex and, in the majority of cases, their structure is unknown. Their application resembles that of vat dyes (see below) [2, 8]. The initially water-insoluble dyes are reduced in the presence of alkali to a water-soluble “leuco” form containing water-solubilizing thiolate ($-S^-$) groups which diffuse into the fiber in the presence of electrolyte. At the end of dyeing, these groups are then oxidized in situ to create a new set of disulfide linkages, producing insoluble colorant which remains trapped in the fiber. Sulfur dyes are widely used for the production of dull deep shades of reasonable wet fastness because of their low cost [10].

Vat Dyes

Members of this class, which are applied mainly to cotton and other cellulosic fibers, generally provide the highest all-round fastness but are expensive, so they tend to be the class of choice when durability is a priority, e.g. production of upholstery [5, 8, 10]. Key aspects of vat dyes are water insolubility and the presence of one or more pairs of carbonyl groups. The class derives its name from the crucial operation during dyeing



Colorant, Textile, Fig. 7 Key features of CI Reactive Blue 15



Colorant, Textile, Fig. 8 Key features of CI Vat Blue 1 (Indigo)

and printing called “vatting” which involves reducing the carbonyl functions of the finely dispersed colorant to produce a water-soluble “leuco” form (see Fig. 8). Once the leuco form of the dye has been applied to the textile, oxidation is then undertaken to regenerate the colored water-insoluble form of the dye within the substrate where it becomes physically trapped. Fastness to washing and light are therefore excellent once loose colorant has been removed by a soaping step. The class is dominated by anthraquinonoid and more complex polycyclic derivatives although indigoid colorants are also important, especially the parent compound Indigo which is the highest volume vat dye (see Fig. 8) [5, 7].

Future Prospects

The enormous variation in both the chemical and physical characteristics of textiles has forced the emergence of a diverse array of coloration technologies. As a consequence, many different types of colorant have been commercialized for application to textiles. Pigments are used in the printing of textiles or in the mass coloration of polymer intended for fiber manufacture, while dyes are applied industrially in different ways ranging from exhaustion dyeing to transfer printing. The wide spectrum of techniques available to a dyer or a printer has only been made possible through the development of colorants that possess chemistries and physical properties tailored for application to specific types of textiles. There is no such thing as

a ‘universal’ colorant that can be applied satisfactorily to all textiles for all intended outlets, nor is one likely to materialize in the near future. It is widely accepted that the chances of new application classes or structural types being introduced are remote. The rise in regulatory costs associated with bringing novel chemical entities to market over the past couple of decades has acted as a brake on the commercialization of new textile colorants. Nevertheless, research along these lines continues alongside incremental work to improve on the technical properties, economy, and environmental impact of existing textile colorants. Despite claims of a renaissance in natural dye usage in textile coloration that continue to appear in the academic literature, such colorants are not a panacea for the disadvantages of synthetic materials owing to unresolved shortcomings associated with manufacture, application, and performance. The fruits of research into structural color, e.g. the design of materials such as cellulosic fibers whose surface structure itself generates color without recourse to use of dyes or pigments, presents a potential existential threat to textile colorants [11]. However, industry will continue to produce and apply synthetic dyes and pigments in the textile coloration sector for many years to come.

Cross-References

- [Coloration, Mordants](#)
- [Colorant, Natural](#)
- [Coloration, Textile](#)

- [Dye](#)
- [Dye, Metal Complex](#)

References

1. Zollinger, H.: Color chemistry: synthesis, properties, and applications of organic dyes and pigments. VCH/VHCA/Wiley-VCH, Zürich/Weinheim (2003)
2. Wardman, R.H.: An introduction to textile coloration: principles and practice. Wiley, Chichester (2018)
3. Towns, A.: Colorants, general survey. Phys. Sci. Rev. **4**(8) (2019). Retrieved 17 August 2020 from <https://doi.org/10.1515/psr-2019-0008>
4. Hofenk de Graaff, J.H. (ed.): The colourful past: origins, chemistry and identification of natural dyestuffs. Abegg-Stiftung and Archetype Publications, Riggisberg/London (2004)
5. Hunger, K. (ed.): Industrial dyes. Wiley-VCH, Weinheim (2003)
6. Herbst, W., Hunger, K.: Industrial organic pigments. Wiley-VCH, Weinheim (1997)
7. Shore, J. (ed.): Colorants and auxiliaries, vol. 1. SDC, Bradford (1990)
8. Waring, D., Hallas, G. (eds.): The chemistry and application of dyes. Plenum, New York (1994)
9. Burkinshaw, S.M.: Physico-chemical aspects of textile coloration. Wiley, Chichester (2016)
10. Richards, P.R.: Dye types and application methods. In: Best, J. (ed.) Colour design: Theories and applications, pp. 471–496. Woodhead Publishing, Cambridge (2012)
11. Mehta, A.: Structural colour. Chem. World. **15**(6) (2018)

Colorant, Thermochromic

- [Colorant, Thermochromic](#)

Colorant, Thermochromic

Mary Anne White and Alex Bourque
Department of Chemistry, Dalhousie University,
Halifax, NS, Canada

Synonyms

[Colorant](#), [halochromic](#); [Colorant](#), [thermo-](#)
[chromatic](#); [Dye](#), [functional](#)

Definition

Thermochromism is the property of temperature dependence of the electronic absorption spectrum of a material, resulting in a color that depends on temperature. At one time, the term thermochromism, also known as thermochromatism, was reserved for isolated compounds and their solutions. However, the advancement of the field has led to a broadening of this definition; the term thermochromism now also describes multicomponent mixtures that are able to change color in response to changes in temperature. For both isolated compounds and mixtures, reversibility of the color change generally is regarded as a necessary condition for thermochromic behavior.

Interestingly, thermal copy and receipt paper, which have been the most commercially important thermally responsive color-changing products for the past several decades, undergo an irreversible coloring reaction and therefore do not fall under this narrow technical definition of thermochromism. High technology products such as thermal copy paper and erasable inks, which make use of multicomponent thermochromic *mixtures*, become thermochromic due to the thermal initiation of a secondary color-changing reaction. *Halochromism*, the color change of a compound in response to changing pH, is the temperature dependent auxiliary process occurring in thermal copy paper which causes the color change.

Introduction

Thermochromic compounds change color reversibly in response to temperature changes [1, 2]. Many thermochromic compounds are organic in nature and undergo thermally activated chemical modifications which give rise to the color change. In many cases, the chemical modification is a result of *tautomerism*. Tautomerism refers to reversible structural isomerism that consists of multiple steps usually involving bond cleavage, molecular reconfiguration, and subsequent bond reformation [3]. Thermochromic behavior can be observed in a wide variety of isolated compounds. Common organic thermochromic compounds include crowded ethenes, Schiff bases, spiro-

heterocycles (e.g., spiropyrans, spironaphthalenes, etc.), and macromolecular systems including liquid crystals and π -conjugated polymeric materials [4]. There are comparatively fewer examples of inorganic thermochromism. However, vanadium (IV) oxide (VO_2) has recently garnered much attention from researchers as a potential smart coating material due to its thermally tunable infrared and near-infrared absorption spectrum [5].

Today, significant research efforts in academia and industry focus on the development of new technologies and devices based on multi-component thermochromic mixtures including smart coatings, erasable printing media, and temperature sensors. This article provides a brief review of important examples of thermochromic compounds, followed by a description of the most recent advancements in the field with an emphasis on applications for new high-technology materials. Advanced *thermochromic materials* take advantage of the growing field of *functional dye chemistry*, and a few examples are presented.

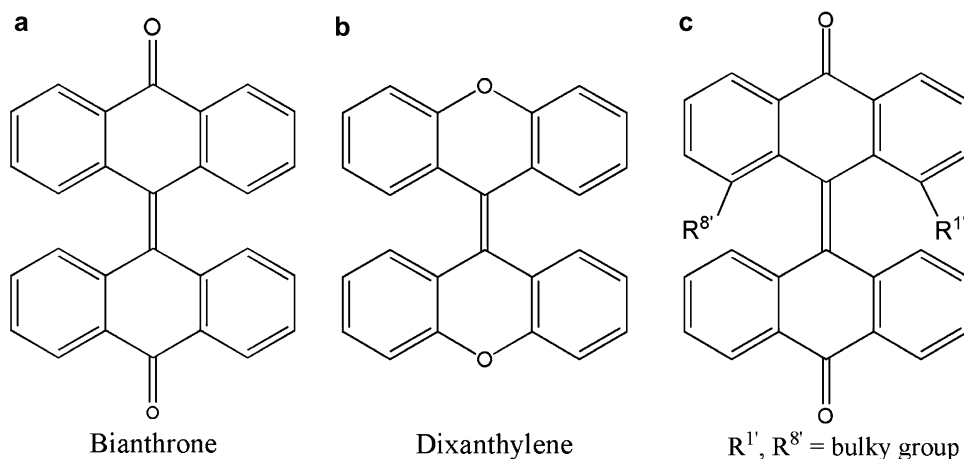
Organic Thermochromism

Thermochromic behavior in organic compounds is often caused by thermally activated chemical rearrangements, i.e., *thermal tautomerism*. Some

of the more common examples of tautomerism include acid-base reactions, keto-enol rearrangement, and lactim-lactam equilibria. Tautomerism is usually influenced by changes in temperature and solvent properties such as composition, polarity, and pH, and thermally activated tautomerism can lead to thermochromism.

Bianthrone and Crowded Ethenes

One of the first examples of reversible thermochromism observed in organic compounds was that of bianthrone [2]. The structures of bianthrone and some of its analogues are shown in Fig. 1. Bianthrone's structure can be considered in terms of the planarity of each anthrone moiety (i.e., the upper and lower half of the molecule, as shown in Fig. 1). At room temperature, the anthrone moieties are curved such that the aromatic rings within each anthrone moiety are not coplanar. When the temperature is increased, the central carbon-carbon double bond expands. This slight bond expansion allows the two anthrone moieties to rotate with respect to one another such that the dihedral angle across the double bond approaches 90° . When this occurs, the previously bent anthrone moieties each take a more planar structure as the steric repulsion created by the proximity of the other anthrone moiety is reduced by rotation about the central bond [6].



Colorant, Thermochromic, Fig. 1 (a) The structure of bianthrone and (b) related compound dixanthylene. (c) Bulky groups at the 1, 1', 8, and/or 8' positions would result in nonthermochromic compounds

The increased planarity at higher temperature permits π -conjugation to extend more effectively across each anthrone moiety, decreasing the HOMO-LUMO gap and concomitant electronic absorption energy, and thereby giving rise to a change in color of the compound with a change in temperature. Bianthrone is yellow in the solid (absorbing violet light at low temperature), and green (absorbing red light) in the melt. Substituents play a strong role in determining if these crowded ethenes will be thermochromic. Dixanthylene is colorless in the solid, and green in the melt. Bulky groups at the 1' and 8' positions would cause excessive steric repulsions and prevent the formation of the bent anthrone state, which is required for the color-changing process [4].

Schiff Bases, Also Known as Salicylidene-Anilines

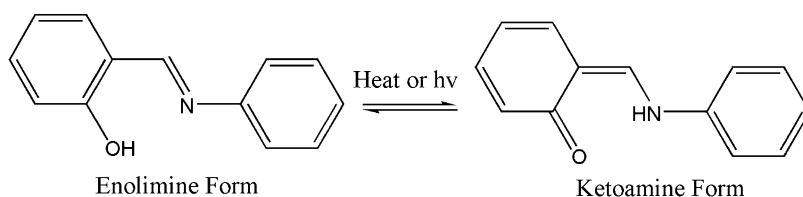
Thermochromic Schiff bases, also known as imines or azomethines, are formed by the condensation of an aromatic amine with either an aldehyde or ketone. The simplest compound in this class, salicylidene-aniline, is generated by the reaction of salicylaldehyde with aniline. The thermochromic Schiff bases exist in a state of equilibrium between two forms, the *enol-imine* and *keto-enamine* forms. For salicylidene-aniline, the enol-imine form is more stable at room temperature (Fig. 2). Increased temperature allows an intramolecular tautomerization to occur: The proton from the hydroxyl oxygen migrates to the imine nitrogen. Due to the rearrangement of the π -electrons, the keto-enamine form has more extended π -bonding than the parent enol-imine, which gives rise to a change in color [7].

Substituent effects play a very important role in this system and define which tautomeric form of the enol-imine—keto-enamine equilibrium dominates. To switch between forms, thermal energy sufficient to exceed the activation energy of the tautomeric reaction must be added to the system. A review of the Schiff bases by Minkin et al. demonstrates the vast variability in this family of compounds, where simple modifications and ring substitutions can push the enol-imine-keto-enamine equilibrium in either direction [8].

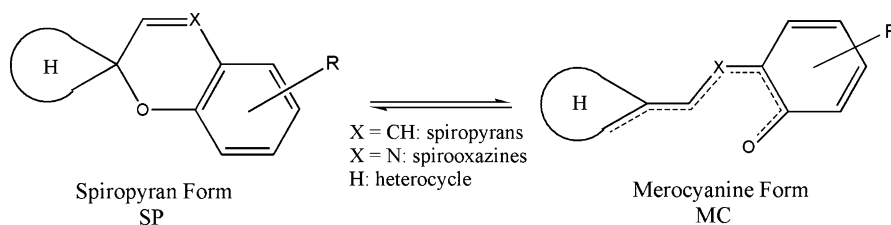
Note that photons (light) also can be used to provide sufficient energy to initiate tautomerism. In that case, the process is considered to be *photochromic*. Thermochromic and photochromic properties in this class of compounds were long thought to be mutually exclusive. However, more recent studies have indicated that salicylideneanilines are almost always thermochromic in the solid state and are occasionally also photochromic [9].

Spiro Compounds, Including Spiropyrans

Spiro compounds are arguably the most important class of compounds used in thermochromic applications; halochromic triarylmethane and fluoran dyes are widely used as colorants in multi-component thermochromic mixtures (for example, in thermal receipt paper and erasable printing media) [10, 11]. This class includes spiropyrans, spironaphthalenes, spiropyrans, and fluoran and triarylmethane dyes. Spiro compounds, so named for the “spiro” central sp^3 tetrahedral carbon center that all members share, are subject to numerous tautomeric equilibria including lactim-lactam, acid-base, and the aforementioned enol-keto equilibria. These equilibria result



Colorant, Thermochromic, Fig. 2 Tautomeric equilibrium between enol-imine and keto-enamine forms observed for thermochromic Schiff bases [8]



Colorant, Thermochromic, Fig. 3 Schematic view of the tautomeric equilibrium between a ring-closed spiropyran (SP) and a ring-opened merocyanine (MC) form as observed for spiro compounds [4]

in chemical modifications that have significant impact on the electronic structure, and subsequently, on the color of these compounds [12]. Many *functional dyes* belong to this category of compounds: Some of the more important examples are the triarylmethane dyes (e.g., crystal violet lactone, CVL) and the fluoran dyes [13].

The spiropyrans form one of the longest-known and best characterized classes in this group. Figure 3 shows an equilibrium that is observed commonly in this class of compounds [4]. The spiro center is located within the pyran ring which, upon excitation with thermal energy, can undergo a ring-opening reaction that alters the electronic structure. As a result of the disruption of the spiropyran (SP) form, the molecule undergoes electronic rearrangement giving the merocyanine (MC) form, which is deeply colored (e.g., violet, red, and blue) for nearly all members of this family [2].

An important structural aspect of this class of compounds is that in the uncolored spiropyran form, two or more aromatic moieties are segregated from each other by the central spiro carbon atom. After the tautomeric rearrangement to the more planar merocyanine form, the formerly segregated π -electronic domains are able to form resonance structures which extend across the entire molecule. This delocalization of the π -electronic structure lowers the HOMO-LUMO gap (which would be in the UV typically for the ring-closed structure) and gives rise to the formation of intense color in these compounds. The mechanism for this process is shown in Fig. 4 for di- β -naphthopyran, which was first studied by Dickinson and Heibron [14]. An important observation is that ring-opening leads to a

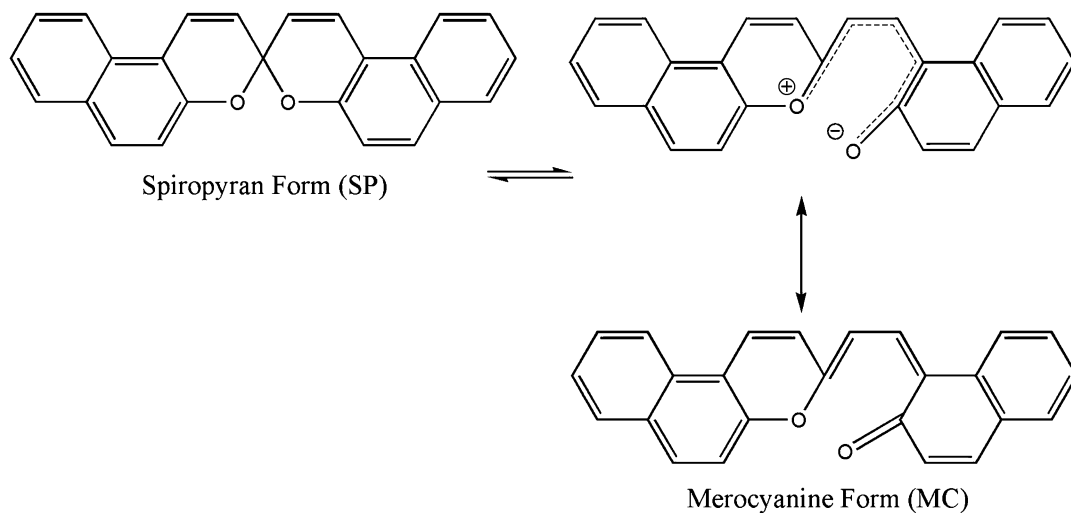
zwitterion, which has significant implications concerning the solubility of the ring-opened merocyanine form.

Ring-opened spiropyrans usually adopt the quinoid structure as shown in the bottom right of Fig. 4. However, if the R-group on the pyran moiety (see Fig. 3) is a strong electron-withdrawing group, the negative charge on the oxygen in the zwitterionic form will be stabilized allowing the molecule to have zwitterionic character. Substitutive modifications to either of the aromatic functionalities in spiropyrans can significantly modify the thermochromic properties by stabilizing the charges formed in the ring-opened configuration.

Liquid Crystals

A well-known example of thermochromic behavior which results from macromolecular interactions is the thermochromic liquid crystals. Thermochromic liquid crystals can be found in products as diverse as mood rings, “stress testers,” warning indicators, and thermometers [15]. Thermochromic effects have been exploited in many different types of liquid crystals, but only two basic series of liquid crystals have found widespread use in thermochromic applications. The esters of cholesterol (Table 1) were first studied by Reinitzer and led to the identification of the liquid crystalline phase [16]. The ester derivatives of (S)-4-(2-methylbutyl) phenol (Table 2) form the basis of the majority of today’s thermochromic liquid crystal devices.

Thermochromic liquid crystals take advantage of the special optical properties of the chiral nematic phase (abbreviated Ch and/or N*). The nematic phase is formed by calamitic (“rod-



Colorant, Thermochromic, Fig. 4 Mechanism of the ring-opening reaction in the spiropyran di- β -naphthopyran. Cleavage of the spiro carbon-oxygen

bond yields a structure with greater planarity and extended π -conjugation. The compound becomes colored due to this extended conjugation [4]

Colorant, Thermochromic, Table 1 Structure and thermochromic temperature ranges of some cholesterol esters [17]

R	Temperature range of color change
CH ₃ CO ₂ to C ₈ H ₁₃ CO ₂	94–113 °C
CH ₃ OCO ₂ to C ₆ H ₁₃ OCO ₂	94–108 °C
C ₇ H ₁₅ OCO ₂ to C ₉ H ₁₉ OCO ₂	77–81 °C

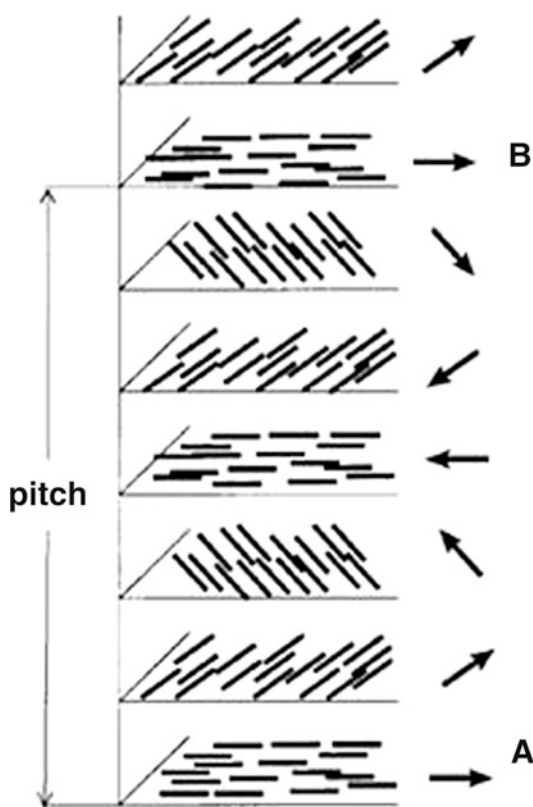
Colorant, Thermochromic, Table 2 Structure and thermochromic temperature ranges of some 2- and 3-ring 2-methylbutyl phenol esters [17]

R	Temperature range of color change
C ₅ H ₁₁ & C ₈ H ₁₇	–5–0 °C
C ₆ H ₁₃ O to C ₁₀ H ₂₁ O	35–50 °C
C ₅ H ₁₁ -Ph to C ₇ H ₁₅ -Ph	130–140 °C

shaped”) liquid crystals and demonstrates only orientational order; this phase lacks long-range positional order [1]. The angular distribution of the long molecular axes of the calamitic molecules is central to the special optical properties of chiral nematic phase liquid crystals. The most probable direction of the long molecular axes defines the *director* which coincides with the principal optical axis of the uniaxial phase. In the chiral nematic phase, the director spirals about a helical axis. The pitch length, p , corresponds to the distance along the helical axis required for the director to make a full rotation about the helical axis. The pitch length can be on the order of a few hundred nanometers, i.e., the wavelength of visible light [1]. This effect is shown schematically in Fig. 5 [18].

Light reflected by the layer in the chiral nematic phase at location A (see Fig. 5) can constructively interfere with light reflected from the layer at position B (see Fig. 5) if the extra distance travelled (layer A compared with layer B) is an integer number of wavelengths of light. This phenomenon is analogous to Bragg reflection in layered crystalline solids. In such a way, chiral nematic phase liquid crystals act as a diffraction grating, or more precisely, a monochromator. Temperature variations in the sample can cause the pitch length to change via thermal expansion, giving rise to variations in the wavelength of light that is constructively reflected (aka selective reflection). An important practical consideration arises from this selective reflection; light that is not reflected by the liquid crystal must be transmitted or absorbed. If the backing material is lightly colored, any transmitted light can be reflected back through the liquid crystal, interfering with the single selected wavelength of reflected light, changing the color. Therefore, thermochromic liquid crystal devices are almost always printed on black backings to absorb light of wavelengths other than the one selected for reflection [17].

To obtain thermochromic behavior in chiral nematic liquid crystals, the pitch must vary rapidly with temperature. Phase transitions are usually exploited to control variation in pitch. The most common transition exploited is the S-N*

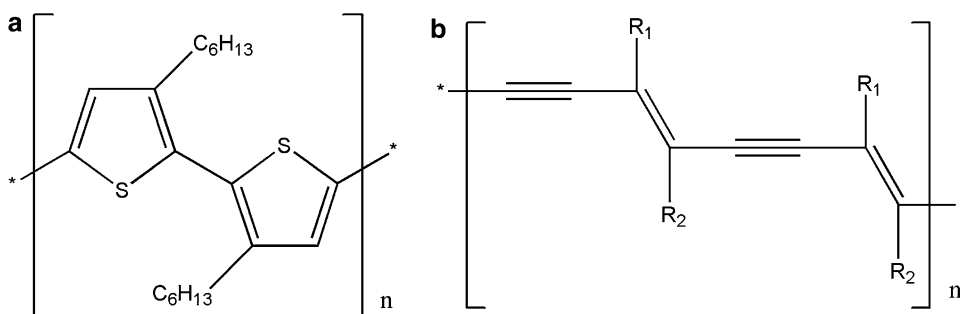
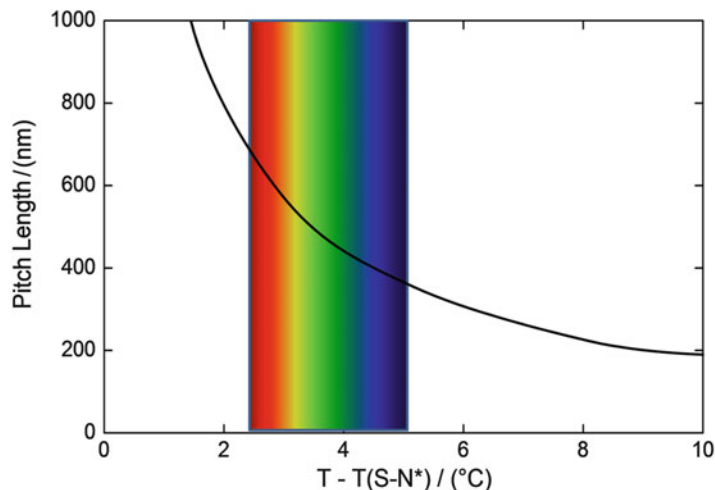


Colorant, Thermochromic, Fig. 5 Pitch length corresponds to the distance required for the director to make a complete rotation about the central helical axis, from layer A to layer B. (Reproduced with permission from *Journal of Chemical Education*, 1999, 76(9), 1201–120. Copyright (1999) American Chemical Society)

(smectic to chiral nematic) phase transition. As the material cools within the N* phase, the pitch elongates (i.e., the liquid crystal becomes more locally structured) and the reflected color changes from blue (at higher temperatures) to red (nearer to the transition temperature). This effect is shown graphically in Fig. 6. Note that further heating of the liquid crystal brings about a transition to the isotropic liquid phase concomitant with a loss of selective reflection and color.

In general, the two important categories of thermochromic liquid crystals behave in much the same way. The major difference can be found in the applicable temperature range for each of the materials. Cholesteric liquid crystals generally have much higher transition temperatures and tend to find applications in

Colorant, Thermochromic, Fig. 6 An illustrative representation of the change in pitch length and corresponding color of a chiral nematic phase liquid crystal above the S-N* transition [17]



Colorant, Thermochromic, Fig. 7 (a) The polymer 3-hexyl polythiophene, an important material for organic electronics, has coplanar thiophene rings at low

temperature. (b) Polydiacetylenes also can display thermochromic behavior depending on the choice of side-chain groups

thermometers on pasteurization equipment, ovens, and warning indicators on hot surfaces. The (S)-4-(2-methylbutyl) phenol derivatives have transitions at much lower temperatures, including physiological temperatures, and find use in thermometers, mood rings, and in refrigerator and food spoilage warning labels. Thermochromic liquid crystal devices can be engineered to behave in both reversible and irreversible manners, with the latter being particularly important if the thermal history of a product (e.g., perishable food products) is of particular importance [15, 19].

Polymers

Highly π -conjugated organic polymers can demonstrate interesting electrical properties resulting from very long conjugation lengths and a high

degree of delocalized electron density. As with conjugated small molecules, conjugated oligomers and polymers are susceptible to thermally induced structural modifications and can exhibit thermochromism. Some common thermochromic polymers include polythiophenes (Fig. 7a), polydiacetylenes (Fig. 7b), and α -conjugated polysilanes [20, 21].

Thermochromism in conjugated polymers arises when sufficient thermal energy causes an order-disorder transition involving the bulky side chains of the polymer. The side-chains in polymers generally keep the polymer backbone organized in some fashion; the backbone bonds tend to be in either all-*trans* or helical conformations. Above a certain temperature, the side chain groups become dynamically disordered and can

no longer keep the backbone chain in its original conformation. The most common transformation is from all-*trans* to *gauche* conformation. Thermochromism results from the change in the HOMO-LUMO gap.

Polythiophenes are widely employed in organic electronics as a conducting layer. They are highly conjugated when the thiophene rings are in a *trans*-planar configuration. Regioregular poly-3-alkylthiophenes (Fig. 7) undergo a reversible color change from red-violet to yellow when heated under vacuum. This change arises from weakening side-chain interactions that are no longer able to maintain the coplanarity of the thiophene rings, resulting in twisting along the chain, a decrease in conjugation, and a change in wavelength of light absorbed [22].

Inorganic Thermochromism

Thermochromism in inorganic materials can have many different origins: changes in ligand geometry, changes in metal coordination, changes in solvation, changes in band-gap energy, changes in reflectance properties, changes in distribution of defects in the material, and phase transitions [23].

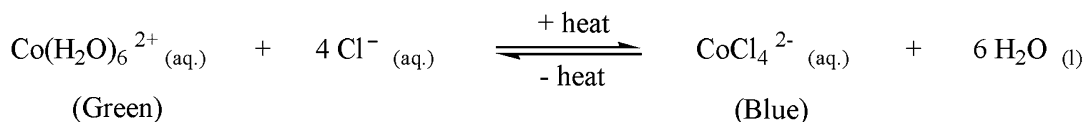
An example of thermochromism arising from a phase transition is the compound Ag_2HgI_4 . At room temperature, the compound adopts a tetragonal crystal structure and is yellow. Upon heating to 50 °C, Ag_2HgI_4 undergoes a first-order phase transition from tetragonal to a cubic phase concomitant with a color change to orange. Upon further heating, it undergoes a gradual (second-order) order-disorder transition to a phase in which the silver ions become mobile in the lattice and the color of the compound changes to black. Therefore, across a temperature range from 25 °C to 75 °C the material changes from yellow to orange to black [24].

Reversible thermochromism also can be observed for inorganic compounds in solution. In solution, the color changes are often associated with modification of the solvation sphere, changes in coordination number, or ligand exchange with the solvent. A commonly cited example is CoCl_2 in water which is blue at room temperature and green at 0 °C, as shown in Fig. 8 [25]. As the system is heated, the Co(II) coordination changes from octahedral to tetrahedral and is coupled with ligand exchange. This structural change alters the electronic field experienced by the central Co^{2+} as a function of temperature, changing the wavelength of light absorbed, and giving rise to thermochromism.

Perhaps the most interesting inorganic thermochromic compound is vanadium (IV) dioxide, VO_2 , which undergoes a semiconductor to metal transition at 68 °C. The transition modifies the absorption spectrum in the infrared and near-infrared regions. Vanadium dioxide is infrared transmissive below ~68 °C and infrared reflecting at higher temperatures [5]. Vanadium dioxide is being considered for use in smart coatings which would allow visible sunlight to pass through a thin film coating but block infrared radiation, thus reducing building cooling requirements [26]. Inorganic thermochromic compounds are of great interest for building coatings owing to their stability to light, which is substantially better than organic thermochromic compounds which are notoriously susceptible to decomposition under extended light exposure [27].

Thermochromic Materials

The term *thermochromic materials* refers to multicomponent mixtures of chemicals which, although not necessarily thermochromic



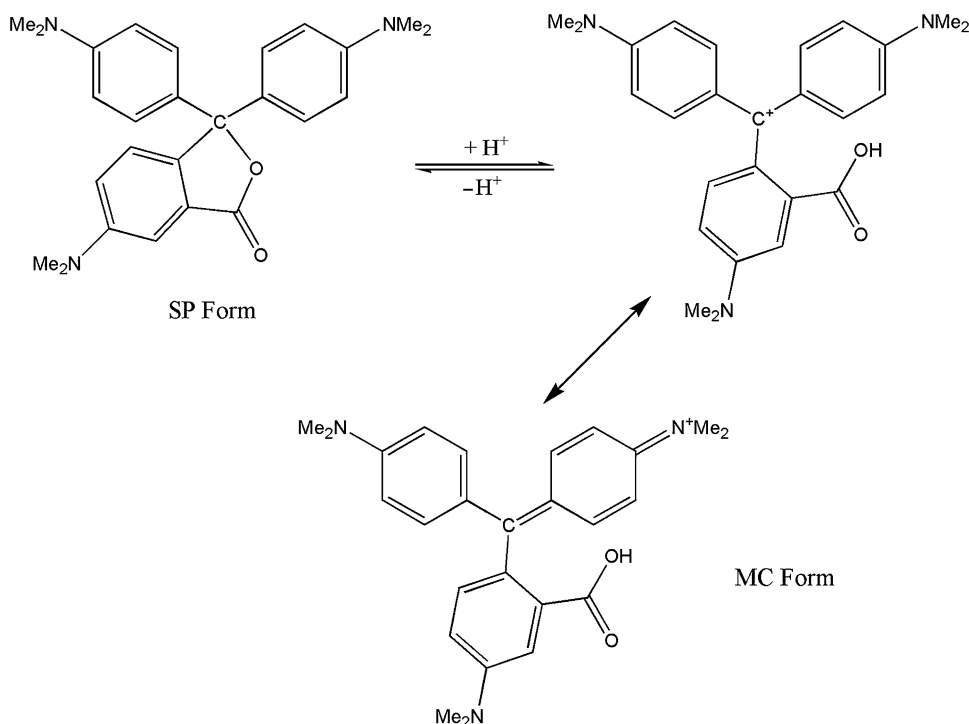
Colorant, Thermochromic, Fig. 8 The equilibrium form for Co^{2+} in aqueous solution and in the presence of Cl^- changes from green to blue upon heating from 0 °C to room temperature, resulting in thermochromism

individually, create a thermochromic system when mixed in the appropriate proportions. Some popular examples of commercial products incorporating this type of thermochromic material are the Pilot Frixion erasable pen, the Coors Light beer bottle label, and the NHL hockey puck [28, 29]. The color of ink from the Frixion pen can be erased thermally by the friction created by rubbing the eraser head on the page [follow this link for a video <https://youtu.be/NbtcVg0m5Q8>] [28]. Similarly, the label on the Coors Light beer bottle contains a color-changing dye system that reversibly changes from colorless to blue upon cooling the container to below $\sim 6^{\circ}\text{C}$, and the NHL hockey puck label changes from violet to colorless to indicate that a puck has warmed and requires replacement [29].

In such mixtures, a color-forming agent, the *chromophore*, also known as *leuco-dye*, reacts with a color-developing agent, the *developer*, to initiate the color-changing reaction. The color-

changing reaction also is controlled by another component of the mixture, usually referred to as the *cosolvent*, which forms the bulk of the mixture. The cosolvent melts, and its melting point determines the color-change temperature. The cosolvent's interactions with the other components also determine if the colored form of the mixture occurs at high or low temperature [30].

Significant effort within industry has been aimed toward development of thermally erasable printing inks and toners for large-scale printing in an office setting [31]. National Cash Register Co. developed an irreversible "thermochromic" receipt paper in the 1950s using the spiroactone dye crystal violet lactone (CVL) as the chromophore [18]. The closed ring SP form of CVL is colorless; upon interaction with an acidic compound (or an electron acceptor), the lactone ring opens forming the intensely blue, ring-opened MC form (Fig. 9). Attaplagus clay and phenolic



Colorant, Thermochromic, Fig. 9 Ring-opening reaction of crystal violet lactone (CVL) in the presence of an acidic developer. The ring-opened charged form has multiple resonance forms, only two of which are shown here

resins were originally used to develop the color, although more recently, phenolic compounds such as bisphenol A (BPA) have been the developers of choice in the United States.

Most of the receipt paper used today in commercial enterprises employs this type of technology, although the chemicals employed are changing. Fluoran dyes are generally used to produce the black color of modern receipt paper and bisphenol A is being replaced by other, less harmful, and phenolic compounds. These compounds include Wincon D8, BASF Pergafast PF201, and to a lesser extent other bisphenols including bisphenol S and bisphenol AF [32].

Today, the color-forming components are printed onto the receipt paper in two layers, one containing the chromophore and the other containing the developer. Both layers are mainly composed of a heat-fusible material called a *sensitizer*. Heating the receipt paper causes the sensitizer to melt, placing the chromophore and developer into intimate contact and thereby initiating the coloring reaction. Although this process is technically not thermochromic due to the lack of reversibility, the widespread use of thermal receipt paper warrants its inclusion in this section [33].

An interesting commercial development from Japan is Toshiba's e-Blue erasable laser jet toner. The toner is composed of a blue colored spirolactone dye and phenolic developer embedded in a polymer matrix. When printed, the toner is blue. Heating a printed page will cause a decolorization reaction in which the developer is segregated from the dye, returning the initial uncolored state of the dye and erasing the printed image [34]. The potential benefits of reducing the amount of paper that enters the recycling waste stream are substantial, although poor resistance to color fade and low image quality thus far have precluded wide usage of such rewritable printing media. These examples of thermochromic materials also fall under the umbrella of the broad *functional dye* field and are discussed further in another chapter in this text.

The previously mentioned Pilot frixion erasable pen ink represents a special case of reversibly thermochromic materials. The cosolvent component of the Pilot Frixion erasable pen demonstrates a large hysteresis between its melting and crystallization temperatures [28]. To maintain the hysteresis effect, the thermochromic material is microencapsulated in a rigid polymeric shell (making a "thermochromic pigment"). The colored form of the thermochromic material occurs when the cosolvent is solid, and the colorless form occurs when the cosolvent is liquid. The large hysteresis of the melting/crystallization temperature of the cosolvent permits both the colored and colorless/erased states to exist at room temperature (the erased state is metastable), allowing for an erasable ink. Extensive research activity in the writing instrument industry is aimed at developing new cosolvent components with large hysteresis. Companies including BIC (Gelocytion Illusion) and Uniball (Fanthom) have also recently developed erasable inks [35, 36].

Thermochromic Colorants

Thermochromic leuco dyes and liquid crystal systems are used in the textile industry for both functional and artistic purposes. The breadth of variation in leuco dye structure permits the formulation of thermochromic products demonstrating an amazing array of colors. The choice of cosolvent allows for precise control of activation temperatures (i.e., the color-changing temperature). Companies supplying thermochromic products can design products to suit the needs of the textile manufacturer. Virtually, any color imaginable can be produced by precise mixing of primary colors (e.g., red, green, blue, etc.), while clever selection of activation temperatures can result in interesting and esthetically pleasing color-play effects [37].

Thermochromic colorants need to be isolated from their surroundings prior to use in textile applications in order to preserve the intended coloring behavior of the colorant system. To this end, microencapsulation is used to isolate the thermochromic system, with the *in situ and interfacial*

polymerization methods being the most common. The microcapsules are dried to form a powder, after which they are usually referred to as *thermochromic pigments*. The pigments can then be made into slurries, emulsions, or pellets, or dispersed into inks and paints, or applied directly to a fabric.

One of the major problems concerning the use of thermochromic pigments is dilution of the dye throughout the processing steps; final dye concentrations can range from 3–5% for pellets to 15–30% for inks and paints [37]. Other problems include poor stability against UV radiation (e.g., photo-bleaching), poor resistance to the effects of water and detergents, the cost of the thermochromic material, and for some, toxicity of the components [27, 32]. In the case of liquid crystals, the fiber onto which the thermochromic pigment is printed must be black to prevent unwanted reflection effects [38]. Additionally, the microencapsulation process can disrupt carefully engineered interactions in multicomponent thermochromic mixtures (i.e., dye, developer, and cosolvent mixtures) such that the final microencapsulated product does not behave in the same way as the isolated system.

The use of thermochromic colorants in the textile industry has been mainly limited to novelty applications (e.g., Hypercolor T-shirts). More recent textile applications have been focused on using the thermochromic effect for artistic purposes [38]. Many of the thermochromic artistic works reviewed by Christie et al. [38] employed fabric-bound, microencapsulated thermochromic dyes coupled with heat-producing microelectronic devices to initiate the color-changing behavior. Smart materials including electronics-coupled textiles [39], multisensory interactive wallpapers [40], surface coatings containing heat-storing *phase change materials* (PCMs), and soft-woven thermochromic fabrics [41] have been reported. These artistic applications of thermochromic colorants demonstrate the important link between scientific and technological development and the creativity of the artistic world.

Cross-References

- [Colorant, Photochromic](#)
- [Dye, Functional](#)
- [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

References

1. White, M.A.: Physical properties of materials, 3rd edn. CRC Press, Boca Raton (2019)
2. Day, J.H.: Thermochromism. *Chem. Rev.* **63**, 65–80 (1963)
3. Christie, R.M.: Colour chemistry. Royal Society of Chemistry, Cambridge (2001)
4. Crano, J.C., Guglielmetti, R.J. (eds.): Organic photochromic and thermochromic compounds, vol. 2. Kluwer Academic/Plenum Publishers, New York (1999)
5. Granqvist, C.G., Lansaker, P.C., Mlyuka, N.R., Niklasson, G.A., Avendano, E.: Progress in chromogenics: new results for electrochromic and thermochromic materials and devices. *Sol. Energy Mater. Sol. Cells.* **93**, 2032–2039 (2009)
6. Sidky, M.M.: Thermochromism. *Chem. Stosow.* **27**, 165–182 (1983)
7. Vitry, J.: Photochromie et Thermochromie. *Chim. et Indus. – Genie Chim.* **102**, 1333–1349 (1969)
8. Minkin, V.I., Tsukanov, A.V., Dubonosov, A.D., Bren, V.A.: Tautomeric Schiff bases: Iono-, solvato-, thermo- and photochromism. *J. Mol. Struct.* **998**, 179–191 (2011)
9. Fujiwara, T., Harada, J., Ogawa, K.: Solid-state thermochromism studied by variable-temperature diffuse reflectance spectroscopy. A new perspective on the chromism of salicylideneanilines. *J. Phys. Chem. B.* **108**, 4035–4038 (2004)
10. US patent US4853256, NCR Corporation
11. US patent US9217941, Toshiba TEC Corp
12. Muthyala, R., Lan, X.: The chemistry of Leuco Triarylmethanes. In: Muthyala, R. (ed.) *Chemistry and applications of Leuco Dyes*, Chap. 5, pp. 125–157. Plenum Press, New York (1997)
13. US patent US9580603, Chromatic Technologies Inc.
14. Dickinson, R.E., Heibron, I.M.: Styrylpyrylium salts. Part IX. Colour Phenomena associated with Benzonaphtha- and Dinaphtha-spiropyranes. *J. Chem. Soc.*, 14–20 (1927)
15. PCT patent application WO2015128655, LCR Hallcrest Ltd.
16. Reinitzer, E., quoted by Lehmann, O. Z.: *Physik. Chemie.* **4**, 462–472 (1889)
17. Sage, I.: Thermochromic liquid crystals. *Liq. Cryst.* **38** (11–12), 1551–1561 (2011)
18. White, M.A., Leblanc, M.: Thermochromism in commercial products. *J. Chem. Ed.* **76**, 1201–1205 (1999)

19. US patent US4468137, National Research Development Corp UK
20. Song, J., Cheng, Q., Stevens, R.C.: Modulating artificial membrane morphology: pH-induced chromatic transition and Nanostructural transformation of a Bolaamphiphilic conjugated polymer from blue helical ribbons to red nanofibers. *J. Am. Chem. Soc.* **123**, 3205–3213 (2001)
21. PCT patent application WO2018220644, Adhikary, R., Shah, D
22. Yang, C., Orfino, F.P., Holdcroft, S.: A phenomenological model for predicting Thermochromism of Regioregular and Nonregioregular poly (3-alkylthiophenes). *Macromolecules.* **29**, 6510–6517 (1996)
23. Day, J.H.: Thermochromism of Inorganic Compounds. *Chem. Rev.* **68**, 649–657 (1968)
24. Schwiertz, J., Geist, A., Eppe, M.: Thermally switchable dispersions of thermochromic Ag₂HgI₄ nanoparticles. *Dalton Trans.*, 2921–2925 (2009)
25. Marinovic, M., Nikolic, R., Savovic, J., Gadzuric, S., Zsigari, I.: Thermochromic complex compounds in phase change materials: possible application in an agricultural greenhouse. *Sol. Energy Mater. Sol. Cells.* **51**, 401–411 (1998)
26. US patent US8709306B2, National Institute of Advanced Industrial Science and Technology AIST, Japan
27. US patent US5879438, Pilot Ink Co. Ltd.
28. US patent US8173052, Pilot Ink Co. Ltd.
29. US patent US9175175, Chromatic Technologies Inc.
30. Bourque, A.N., White, M.A.: Control of thermochromic behaviour in crystal violet lactone (CVL)/alkyl gallate/alcohol ternary mixtures. *Can. J. Chem.* **93**(1), 22–31 (2015)
31. US patent US6203603, Kabushiki Kaisha Toshiba
32. Björnsdotter, M.K., Jonker, W., Legradi, J., Kool, J., Ballesteros-Gómez, A.: Bisphenol A alternatives in thermal paper from the Netherlands, Spain, Sweden and Norway. Screening and potential toxicity. *Sci. Total Environ.* **601–602**, 210–221 (2017)
33. US patent US4507670, Ricoh Company, Ltd.
34. Sekiguchi, Y., Takayama, S., Gotanda, T., Sano, K.: Molecular structures of the coloring species of a Leuco dye with phenolic color developers. *Chem. Lett.* **36**, 1010–1011 (2007)
35. PCT patent application, WO2018150146, Societe BIC
36. European patent EP3042942, Mitsubishi Pencil Co.
37. <http://www.hwsands.com/>
38. Christie, R.M., Robertson, S., Taylor, S.: Design concepts for a temperature-sensitive environment using thermochromic colour change. *Colour. Des. Creat.* **1**(1), 1–11 (2007)
39. Worbin, L.: *Nordic Text. J.* 51–69 (2005)
40. US Patent, US7632554 B2, Dong-A University Research Foundation for Industry Academy Cooperation
41. Xslabs. <http://www.xslabs.net/work-pages/krakow.html>

Coloration, Fastness

Martin Bide

Department of Textiles, University of Rhode Island, Fashion Merchandising and Design, Kingston, RI, USA

Definition

Humans use a range of technologies to produce color: solid surfaces are painted, paper and packaging is printed, color images are viewed on screens, and they color the textiles and leather used in their clothing and around homes. Screen color is transient, and few paper products are expected to last long. More durable color is found on painted surfaces and textiles, and the color is provided to them by colorants. That durability is expressed as “fastness.”

Colorants, with the minor exception of interference pigments, have chemical structures that efficiently absorb light and are thus strongly colored. Present in small amounts, they provide their color to an item. That color may change for a number of reasons. Some transient changes can occur, such as when an item becomes wet, or if the colorant undergoes reversible structural changes under the influence of heat (thermochromism) or light (photochromism). Beyond that, if a colored item maintains its color, it is said to possess fastness. Otherwise, the color of the item may change when either:

- The colorant remains in place but is permanently changed or destroyed.
- The colorant is unaltered, and maintains its color, but is removed, for example, by washing.

Such a change or loss of color is described as a *lack* of fastness. If the colorant is removed, for example, during laundering, it can stain other items with which it comes into contact. Even when the color of the original item is not appreciably altered, this is also considered a lack of fastness.

Color Fastness May Thus Be Defined as the Resistance to Change or Removal of the Color of an Item Under End-Use Conditions

Colorants are divided into dyes and pigments. The difference between them affects both their application and the likelihood of their permanence (Colorant, Classification).

A pigment consists of molecular aggregates, usually in the micron range, with minimal solubility in any solvent with which it is likely to come into contact with (► [Pigment \(Organic and Inorganic\)](#)). Its removal by dissolution is therefore unlikely. If a few of the pigment atoms or molecules on the surface of an aggregate particle are removed or changed, the overall color of the particle (and hence of the substrate it is coloring) is little affected. Low solubility and particulate nature thus mean that fastness in coloration by pigments is readily achieved.

In contrast, a dye is characterized by its solubility, most often in water, in its application (► [Dye](#)). In many cases its resistance to removal derives from the forces that drive its absorption by a substrate, and it remains soluble when the colored item is in use. The details are discussed in greater depth below, but the solubility may limit the resistance of the colorant to removal, and the monomolecular dispersion makes the effect of any removal or destruction more apparent.

Dyes are little used for non-textile items. Pigments have some use on textiles, but dyes and textiles are largely mutually inclusive. Solid surfaces typically do little flexing or moving, whereas textiles face a greater range of challenges than most other colored objects: as clothing they must have sufficient flexibility to move with the wearer and be soft enough not to irritate. They rub against each other. Curtains hang in sunny windows. Textiles are regularly cleaned and often must withstand hot water and detergent or the solvents used in dry-cleaning. For these reasons, the subject of color fastness is most widely studied for textiles. As discussed below, truly permanent color is an unrealistic expectation, and the extent to which color change is acceptable varies widely and is influenced by the cost of the item, the cost of replacing it (or its color), and the possible effect of the color loss on other items. Problems of

fastness are therefore strongly associated with dyed textiles, and fastness in other colored materials is often tested using methods and principles drawn from textile tests.

Overview

As outlined in the definition, the color of an item may change when either the colorant remains in place but is permanently changed or destroyed, or when the colorant is unaltered and maintains its color, but is removed. These two cases are considered below.

Color Change: Colorant Destruction

The colorant on an item may undergo a chemical reaction that changes or destroys its color. Such a change is perceived as a lack of fastness. The reaction may take place slowly and only become apparent after a long period of time or may be quite rapid.

The former relates most often to the chemical reactions that occur as a result of exposure to light. Absorption of light raises a colorant molecule to an excited state in which condition it is susceptible to decomposition, or reaction with oxygen, water, or other available reactants. The reactions involved are complex. They have been the subject of extensive study but much remains to be fully understood: the factors that affect the reaction are many and varied (light intensity, spectral power distribution, temperature, humidity, etc.); it is also clear that color change is not simply a property of the chemical identity of the colorant alone, but is affected by colorant-substrate interactions, by colorant aggregation, and (in mixtures) by colorant-colorant interactions [1, 2]. The same dye can have different fastness properties on different fiber types. Pigments are less involved in these external interactions, but their physical form (particle size, crystalline form, etc.) may affect their fastness.

Whether dye or pigment, for a constant exposure, the destruction of colorant occurs at a fairly constant rate. The effect on the color, however, is not constant. A darker color, resulting from a greater amount of colorant, will be affected less

by a given exposure than a light one, where the same amount of colorant destroyed represents a greater proportion of the colorant present. In other words, a pale shade of a given colorant will have lower fastness to colorant destruction. This is in contrast to the lack of fastness represented by staining as a result of colorant removal discussed below.

The more rapid destruction of dye is usually the result of exposure to a chemical agent: most obviously this might be sodium hypochlorite or hydrogen peroxide used as a bleach, but other chemicals such as benzoyl peroxide used in skin medication will effectively decolorize many dyes. Again, pigments are less prone to such reactions, given their particulate nature and insolubility.

Color Change and Staining: Colorant Removal

Color may be removed from an item, particularly a textile, for a number of reasons. Unlike the case of colorant destruction, the color remains, and the removed colorant may color an item with which it comes into contact: such staining is a component of (a lack of) fastness. For reasons that will become clear, once again, this is best examined from the point of view of dyes.

Dyes are applied to textiles in a number of ways: batch dyeing, continuous dyeing, and textile printing (► [Coloration, Textile](#)). The principles of the interaction are common to all, but best illustrated in the case of batch dyeing [3, 4]. In a batch dyeing process, dissolved dye molecules are preferentially sorbed from the external dye solution at the fiber surface and penetrate the interior of the fiber as a result of the intermolecular forces that comprise the “substantivity” of the dye. The relative motion of the dyebath and the substrate replenishes the dye-depleted solution at the fiber surface, and the rate-determining step of the overall process is typically the rate of sorption into the fiber. If dyeing conditions are maintained long enough, an equilibrium is established between dye in fiber and dye in solution, and all fibers are fully and equally penetrated. In practice this situation is rarely achieved, and microscopic observed examination of a dyed textile will reveal fibers and yarns with only partial penetration of dye. Such microscopic unlevelness is tolerable.

An essential requirement of a colored textile is that the color be level at the macroscopic scale. Assuming a clean and absorbent substrate from earlier preparation processes, and homogeneously dyeable fibers, levelness results from a balance of two possible routes. Either the dye is absorbed in a level manner, or an initial unlevel sorption is made level by the migration of dye [5].

The substantivity of the dye is based on the molecular size and shape of the dye and functional groups present. Together, these allow the formation of a range of non-covalent bonding that makes up the attraction of dye for fiber and which enables dyeing to take place [4]. If nothing else happens, these same forces will provide the fastness the dye has in use, in competition with the solubility of the dye in the particular medium (e.g., laundry liquor).

For fastness (as resistance to removal), a high substantivity (strong dye-fiber bonding) is required. But dyes with high substantivity do not migrate readily, and if initial application is unlevel, it is difficult to achieve levelness via migration. To use a high-substantivity dye, the dyer must control the initial uptake (“strike”) of the dye to be as level as possible. For a given system (machine, dye, fiber), dye uptake can be controlled by control of temperature (rate of rise and ultimate temperature), time, and dyebath conditions of ionic strength (salt) pH, and auxiliaries that moderate the dye-fiber interactions (often surfactant based). However, the demands of levelness may require that lower- substantivity (and thus less fast) dyes be used.

As discussed earlier with colorant destruction, the fastness of a particular dyeing will vary with the depth of shade. For the case of colorant removal, a dark shade will exhibit lower fastness than a pale one: the same percentage of colorant removed from a dark shade will have the propensity to stain more heavily than from a pale shade.

For some dye-fiber systems, the substantivity of the dye that drives dyeing is all that provides fastness in later use. Most notably, acid and metal-complex dyes on protein and polyamide fibers (wool, silk, and nylon) direct dyes on cellulosic fibers and disperse dyes on acetate. In essence, a subsequent aqueous challenge represents the

equivalent of a dyebath and the (beginnings of) reestablishment of an equilibrium in which some proportion of the dye is lost from the substrate.

Several dye-fiber systems have fastness attributable to reasons beyond those provided simply by the dye-fiber interactions. Disperse dyes on polyester have fastness greater than the same dyes on acetate. The high glass transition temperature (T_g) of polyester means that practical dyeing takes place only at temperatures considerably higher than that of boiling water: these are found in pressure dyeing machinery. After dyeing (and in use), the fiber is cooled below T_g , and the dye is essentially trapped in the fiber: the inaccessibility of the dye to outside agencies is reflected in the use of “reduction clearing” to remove any surface deposits of the low-solubility dye. Similar arguments apply to the fastness of basic dyes applied to acrylic fibers.

While a dye must be soluble during application, greater fastness can be achieved if the dye is later converted (or reverts) to an insoluble or less soluble form. This form of the dye may also comprise molecular aggregates and/or larger molecules for which physical entrapment may contribute to the fastness. Vat dyes for cellulosic fibers are produced and sold in their insoluble form and must be chemically reduced to a soluble and substantive form for application. Once in the fiber, they are oxidized to their original insoluble form and forms aggregates inside the fiber: essentially they are now pigments. Several dye types are formed inside the fiber as low-solubility moieties as part of their application. Sulfur dyes for cellulosic fibers are supplied in reduced form; and once applied, they undergo a similar oxidation to vat dyes to form a low-solubility dye. Azoic colorants are generated by reaction within the fiber of a soluble coupling component and a soluble diazo component. The resulting azo dye is insoluble and, like a vat dye, aggregates into pigment form. Mordant (“chrome”) dyes on wool rely on a reaction of dye and mordant in situ within the fiber to form a low-solubility complex of good fastness, and its fastness is enhanced by the entrapment of a physically large molecule.

A further reason for additional fastness may arise from the presence in the dye of functional

groups capable of reaction with the substrate to form covalent bonds. So-called reactive dyes achieve their fastness in this way and since their introduction in 1956 have become the dominant dye type for cellulosic fibers, with additional usefulness for wool dyeing where they can replace the mordant dyes and their associated use of heavy metals [6]. It should be noted that the reaction efficiency is not 100%, and at the end of the dyeing process, any unfixed dye that is not removed by washing will be held only by the forces of substantivity and may be readily removed in later use. In such a case the dyeing as a whole will be considered to have poor fastness.

Achieving Fastness

Colorfastness, whether as resistance to destruction or resistance to removal, is largely achieved by choice of colorant. Dye manufacturers will supply customers with “pattern cards” of materials colored with individual dyes, together with the results of key fastness tests (see below) conducted on those materials. Colorists may thus select colorants that are likely to meet fastness requirements, although as mentioned earlier, the depth of shade will affect the fastness achieved. Resistance to removal may additionally depend on the ability of the process to eliminate any dye that is not fixed or that is loosely held on the fiber surface.

Pigments are used for textiles in one of two very different ways. A pigment can be added to the polymer melt or solution used to form manufactured fibers in the same way that a pigment is added to a melt used to mold plastic items. Subsequent solidification traps the pigment within the polymeric matrix. Fibers colored this way are among the most colorfast of all. However, for textiles, the time between fiber manufacture and ultimate sale may be a year or more, and the holding of multiple color stocks of fiber and yarn means that this coloration method is of limited use in this context.

Pigments can also be bound to a textile material by use of a polymeric binder: this is akin to the

application of paint to a surface. The fastness of such colored materials is related to both the colorant and the binder and the strength of the pigment-binder-fiber interactions [7]. Once again, textiles tend to face greater challenges as they flex and move.

Fastness Testing

The permanence of color in an item is desirable, but the challenges it might face in use are many and varied. It is unrealistic and expensive to demand the best fastness in every case, so compromises are inevitably made. The questions then become, what are the likely challenges this item will face? How well *should* it be able to resist them? How well *does* it resist them? The answer to the first derives directly from the intended end use, while the answers to the second and third require some way of presenting the challenge and assessing the resistance. The most realistic way to do this is to conduct a trial with the item in its intended use to see when and how failure occurs. Such real-life trials are occasionally performed, but they are very expensive and take a long time to produce results. Standardized laboratory fastness testing provides an economical and useful alternative. Standard test methods are developed that are designed to approximate to a single real-life challenge and to predict how an item will respond. Such tests typically provide results rapidly, either because the challenge itself is a rapid one (e.g., color being rubbed off) or because the test accelerates the challenge (e.g., using a very intense light instead of real daylight).

A fastness test is developed to provide the challenge and a way of assessing the result. The result is then compared to a requirement that forms part of an overall product specification. The tests may be widely applicable or represent a challenge that is more specific. Tests are developed by standard setting organizations. Much of the original work to develop these tests for textiles was carried out by the Society of Dyers and Colourists (SDC) [8] in the UK and the American Association of Textile Chemists and Colorists

(AATCC) in the USA [9]. With the rise in global trade, fastness tests have become more international, and many countries now rely on tests published under the auspices of the International Organization for Standardization (ISO) [10].

A good test should be valid: the property it measures should have some real-life relevance, and the test should predict in-service suitability. Additionally, a good test should be simple in terms of how it is performed and how easy the instructions are to understand. It should also be reproducible giving the same results from operator to operator and laboratory to laboratory. In some cases, the control of test conditions may be difficult, and standard fabrics of known susceptibility are used as control materials.

Issues of environmental, health, and sustainability have come to the fore in recent years. Several certification schemes are intended to reassure the consumer that an item has been produced in an environmentally friendly manner and that it does not contain (or will release) substances of concern. Colorants are among such substances, and the measure of colorant release is often based on standard color fastness testing. Thus, for example, a range of color fastness tests is included in the OEKO-TEX 100 certification [11] and in the Global Organic Textile Standard [12].

Testing for Fastness: Challenges

As discussed earlier, fastness tests are designed to reproduce the challenges that a colored item would face in real life. Again, the subject tends to emphasize textiles. Some tests relate to challenges faced in the sequence of manufacturing steps that a textile item undergoes before it gets to the consumer. Most relate to the challenges encountered after the item is sold: these can be divided into those met when the item is being used and those involved in its cleaning or refurbishment. Full details of the tests are sold by from the relevant standard setting organizations. The titles and scope are available in texts and on the organizations' web sites [13–15].

Fastness to (Textile) Production Processes

Textiles woven from multicolored yarns may be scoured and may include a white portion that requires bleaching, and thus a test for fastness to bleaching may be needed. Mercerizing can alter color. Dry heat can cause color changes. Wool undergoes a variety of wet treatments: tests determine the color changes caused when wool is carbonized, boiled in water, bleached, set with steam, or milled. The fastness challenges of the last are reflected in the naming of certain acid dyes that will withstand this process as “milling acid dyes.”

After makeup, garments may be pleated or hot-pressed: these can cause thermochromic or sublimation-based color changes. Silk fabrics may be degummed in hot alkaline soap solution, and a fastness test can predict any changes to colored materials that undergo this process.

Fastness in Use

In use, garments may be rubbed, and their color transferred to another item: fastness to rubbing (also known as “crocking”) is one of the most common tests.

Spots of liquid can cause a color change, either by moving dye within the textile or removing the dye and transferring it to an adjacent material. Thus there are tests for spotting by water, dry-cleaning solvent, seawater, perspiration, acids, and alkalis. Swimming pool water with chlorine represents a destructive challenge.

Atmospheric contaminants, notably ozone and oxides of nitrogen (“burnt gas fumes”) derived from combustion and sunlight, can destroy colorants.

Light represents a particular challenge. Real-life exposure is both slow to produce change and is highly variable (but such real-life exposure is included among standard tests). More usually, an accelerated test uses a xenon arc lamp and controls the temperature and humidity of exposure. The test may be extended by the inclusion of water sprays to cover “weathering.”

Fastness to Refurbishment

Textile items get dirty and require periodic cleaning. The processes to be used are given in a care label, and the suggested cleaning method should have been shown by testing to be appropriate to restore the item to an acceptable level while not causing any damage. However, such refurbishment represents another set of challenges to the color.

Laundrying is conducted in aqueous surfactant solutions with agitation. The details can vary considerably: the detergent (type and amount used), the temperature, the amount of water and the extent of agitation (based on the type of machine used), the nature of the materials used to make up the bulk of the load (“ballast”), and the presence or absence of bleach. The tests to assess fastness reflect this variation in practice. Full-scale laundryings, repeated three or five times, may be used, but since each item should be tested individually to avoid possible confusion in results, an accelerated test is used. A small sample, usually with a standard multifiber adjacent fabric, is agitated in standard detergent solution in a small cylinder for 45 min or so. The effect on color will approximate to that obtained in five real cycles of laundrying. Similar considerations apply to dry-cleaning.

Ironing can affect color: tests examine the effects of dry, moist, and wet heat.

Assessing the Results of Fastness Tests

The tests produce color changes and thus assessment of the results is an assessment of color difference [16]. This might be the difference between the original color and the challenged color or the difference between an unstained fabric and one stained by color removed from the test sample. Color difference assessment has been widely studied for judging colors for acceptable closeness to a standard color (“a match”). That use concentrates on small color differences and is as much concerned with the quality of difference as the quantity of difference. However, the assessment of color differences from fastness testing covers a much wider (quantity) range of color differences and is rarely

Coloration, Fastness,

Fig. 1 Coloration fastness:
AATCC gray scale for color
change. (Copyright
AATCC)



concerned with the quality of difference. These two aspects of color difference measurement have thus tended to go their separate ways.

The results from a fastness test can be assessed subjectively, by reference to physical color standards ("gray scales") or objectively, based on spectrophotometric or digital camera measurements. Physical gray scales have been so extensively used that even when an instrument is used, its output is converted to gray scale ratings.

Gray Scales

The use of physical standard color difference pairs has long been standard practice in fastness testing. These physical standards include a range of difference pairs, from small to quite large. There are two gray scales. One is used to assess a change of color,

and the second is used to assess the extent to which a white material is stained. The gray scale for color change (Fig. 1) consists of pairs of neutral gray chips, one of the pair being a constant gray color, with the second ranging from the same to very different arranged in order of increasing difference. The pair with no difference is labeled "5," and the increasingly different pairs are 4.5, 4, 3.5, 3, 2.5, 2, 1.5, and 1.

The gray scale for staining is similar, but the constant color is white, and the increasing difference is provided by successively darker grays (Fig. 2). In addition to the two basic gray scales, for color change and for staining, AATCC has produced a "chromatic transference scale" (Fig. 3). This is used mainly in the assessment of the staining produced in rubbing fastness tests.

The correct use of each of these scales is described in the relevant ISO and AATCC



Coloration, Fastness, Fig. 2 Coloration fastness: AATCC gray scale for staining. (Copyright AATCC)

documents. The angle of viewing, the quality and intensity of the illumination, and the masking of the test specimen/gray scale pair are specified to eliminate variables and increase interlaboratory agreement.

Instrumental Color Measurement of Fastness Test Data

A numerical color difference equation might reproducibly express the color change or staining resulting from a fastness test. The need for this to correspond to human visual data over a wide color range and to make it agree with the results generated visually with gray scales has made this somewhat challenging. Nonetheless, while some deficiencies are recognized, mathematical

formula for gray scale ratings derived from spectrophotometric measurements or digital camera data are published and used. Work to develop better formulae continues.

Conclusions

Color is an important property of many items, and its durability in items that are expected to last over a period of time is a matter of concern. Textile colorants have the greatest potential to produce unwanted changes in the textiles they color and to stain adjacent materials, and textiles are subject to the greatest range of potentially damaging agents. Thus color fastness testing is most widely studied and performed on textiles. The tests reflect real-life challenges encountered in textile processing,



Coloration, Fastness, Fig. 3 Coloration fastness: AATCC chromatic transference scale. (Copyright AATCC)

in use, and in care, and many of them are accelerated. ISO and AATCC are the most active organizations in the development of such tests.

The tests require a measurement of the changes produced by the test challenge and the stains caused by removed color. For many years, measurement has been based on visual comparison with standard gray scales. Compared to other situations involving color assessment, the adoption of instrumental methods to assess fastness results has been slow. Standard methods for such instrumental assessments are in place, if yet to be widely adopted.

Cross-References

- [Coloration, Textile](#)
- [Dye](#)
- [Inorganic, Hybrid, and Functional Pigments](#)

References

1. Haillant, O.: The photofading of colored materials. *Sunspots*. **39**(84), 1–11 (2009). www.atlas-mts.com/fileadmin/downloads/SunSpots/vol39issue84.pdf. Accessed 20 Jan 2013
2. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*, pp. 245–280. VCH, Weinheim (1987)
3. Broadbent, A.: An introduction to dyes and dyeing, and dyeing theory. In: Broadbent, A. (ed.) *Basic Principles of Textile Coloration*, pp. 174–214. Society of Dyers and Colourists, Bradford (2001)
4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Etters, J.N.: Linear exhaustion versus migration: which is the better road to Rome? *J. Soc. Dy. Col.* **112**(3), 75–80 (1996)
6. Renfrew, H.: *Reactive Dyes for Textile Fibers*. Society of Dyers and Colourists, Bradford (1999)
7. Anon: Factors Affecting the Drycleaning Fastness of Pigment Prints. In: *Pigment printing handbook*, pp. 131–142. American Association of Textile Chemists and Colorists, Research Triangle Park (1995)
8. The Society of Dyers and Colourists: <http://www.sdc.org.uk/>. Accessed 5 Feb 2013

9. The American Association of Textile Chemists and Colorists: www.aatcc.org. Accessed 5 Feb 2013
10. Smith, P.: Colour fastness testing methods and equipment. *Rev. Progress. Color.* **24**, 31–40 (1994)
11. Oekotex standard 100: Limit values and fastness. <https://www.oeko-tex.com/en/our-standards/standard-100-by-oeko-tex>. Accessed 16 Apr 2021
12. Global Organic Textile Standard: Section 2.4.14, Technical quality parameters. www.global-standard.org/the-standard/latestversion.html_gots-version3_01march2011.pdf p18. Accessed 21 Jan 2013
13. ISO Standards Catalog: TC 38/SC1 – tests for colored textiles and colorants. <https://webstore.ansi.org/Standards/PCC/ISOCatalog>. Accessed 16 Apr 2021
14. AATCC Test Methods & Evaluation Procedures: www.aatcc.org/testing/methods/index.htm. Accessed 21 Jan 2013
15. Bide, M.: Testing textile durability. In: Annis, P. (ed.) *Understanding and Improving the Durability of Textiles*, pp. 126–142. Woodhead, Cambridge (2012)
16. Bide, M.: Colour measurement and fastness assessment. In: Gulrajani, M. (ed.) *Colour Measurement: Principles, Advances and Industrial Applications*, pp. 196–217. Cambridge, Woodhead (2010)

Coloration, Hair

Tova N. Williams
Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Synonyms

[Environmental chemistry of synthetic dyes; Hair dyes](#)

Definition

The coloration of hair is a long-standing practice that can be traced back to the Egyptians who used Henna to impart color to their hair [1]. Today, hair dyes are used by billions of individuals dissatisfied with the natural, and in many cases, graying of their hair. Many temporary, semipermanent, and permanent hair dyes – all synthetic in nature – are commercially available, to offer the consumer diversity in color options and performance levels. Permanent hair dyes are the most popular dyes,

because they offer the consumer the best degree of permanence to washing and gray coverage. Permanent hair dyes are formed inside the hair fiber via an oxidation of precursors that results in the formation of oligomeric dyes. These dyes are too large to be desorbed from the hair fiber during the washing process. Over the last few decades, permanent hair dyes, in particular the precursors (e.g., *p*-phenylenediamine or PPD) used to develop them in hair, have been scrutinized for their genotoxic and skin sensitization potential. Thus, efforts to design environmentally friendly precursors and coloration technologies have been investigated in recent years.

Overview of the Dyeing of Human Hair

Despite the thousands of dyes that can display substantivity toward keratin-based fibers, only a few are suitable for the commercial coloration of hair [2]. A key limitation is the temperature of dyeing. Though keratinic fibers such as wool can appreciably swell in water at temperatures exceeding 75 °C, the human scalp cannot withstand such high temperatures [2]. Consequently, dyeing is performed at a maximum temperature of 40 °C [3–6]. Dyeing times vary, but generally do not exceed 40 min [2]. To penetrate the hair, the dyes must be soluble in water, not too large, not strongly aggregated in solution, and contain few anionic groups [2]. Also, the dyes should be stable to air, light, friction, perspiration, chlorine, other cosmetic treatments, and the pH at which dyeing is performed [3, 5]. Perhaps of more importance to the consumer, the dyes must be appealing and removable, if desired [2], and impart minimal damage during application [5]. The dyes should also be toxicologically safe [2–5], not only for the consumer but also for the hair stylist who applies the dye [5].

Compared to the synthetic hair dye coloration process, coloration using natural hair dyes is often longer and requires more dye to achieve a satisfactory shade depth [7]. Also, the resultant hair color is typically unstable to light and shampoo treatments [8]. Thus, synthetic dyes have dominated the global hair dye market since the

development of the first dye [7, 9]. Synthetic hair dyes can be categorized as temporary, semipermanent, and permanent. As their names imply, these classes are based on the resistance of the dyes to removal during shampooing or their “permanence.” In this regard, temporary dyes last through only one or two shampoos, while semipermanent dyes last up to six shampoos [4, 5, 10, 11]. Permanent dyes essentially last until new hair growth appears, which is every 4–6 weeks for the average consumer [3, 11, 12]. Dyes that are less permanent tend to be larger and adsorb onto the surface of hair [13] or diffuse only into the cuticle, or the outermost region of the hair fiber [11]. The precursors that give rise to permanent hair dyes are capable of diffusing into the cortex (an inner region of the hair fiber), due to their small size, where they react to form oligomers that aggregate and become too large to be desorbed from the hair [2, 4, 5, 12]. In each case, some dye can remain adsorbed onto the surface of hair [11, 13]. Dye adsorbed onto the surface or within the outer layers of hair can more easily be removed upon shampoo treatment [1, 4, 5, 9–11], especially if dye is weakly bound and the cuticle is damaged (more porous).

Synthetic hair dyes can also be classified based on their chemistry as non-oxidative or oxidative, the latter of which describes the permanent hair dye formation process that involves the oxidative coupling of precursors to produce the dyes within hair. Temporary and semipermanent dyes are

designated as non-oxidative dyes, because they do not require an oxidant for color formation [1, 9]. Temporary dyes are ionic in nature, while semipermanent dyes are either ionic or nonionic. Examples of temporary and semipermanent dyes can be found in Hair Dye Substance Database recently developed by Williams and co-workers [14], which is available for download at https://figshare.com/articles/Hair_Dye_Substance_Database_HDSD_/5505856. Each of the dye classes, with respect to their diffusion into hair after dyeing, is schematically represented in Fig. 1.

Hair Dye Diffusion

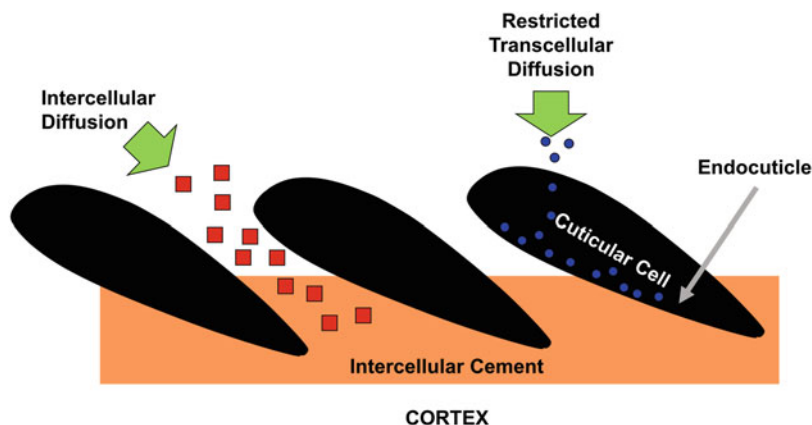
The principle mechanism of dye diffusion in hair has been a controversial topic [15–19]. Two theories are recognized – intercellular and restricted transcellular (see Fig. 2), and it has been suggested that they coexist [20, 21]. A key difference between the two theories is the location at which the dye first penetrates the fiber. In today’s more accepted theory [19, 20], intercellular diffusion, dye penetrates the fiber initially through the intercellular cement or between cuticular cells parallel to the hair fiber [15, 16, 22–24]. Dye diffuses into the regions low in sulfur content and then into the regions high in sulfur content [15, 16, 22–24]. The intermacrofibrillar cement in cortical cells can also provide a diffusion path



Coloration, Hair, Fig. 1 Schematic representation of dye diffusion into hair fiber cross-sections based on dye type

Coloration, Hair,

Fig. 2 Schematic representation of the intercellular and restricted transcellular hair dye diffusion mechanisms. (Note: the endocuticle spans the length of the cuticular cell; large, hydrophobic dye molecules are shown in red “squares”, and small, hydrophilic dye molecules are shown in blue “circles”)



[24]. Contrarily, restricted transcellular diffusion occurs via the penetration of dye normal to the hair fiber through the cuticular cell edge, namely, through the endocuticle and cell membrane complex (CMC) [19]. The preferred mechanism for large, hydrophobic or apolar molecules is suggested to be intercellular diffusion, while the preferred mechanism for small, hydrophilic molecules is suggested to be transcellular diffusion [25].

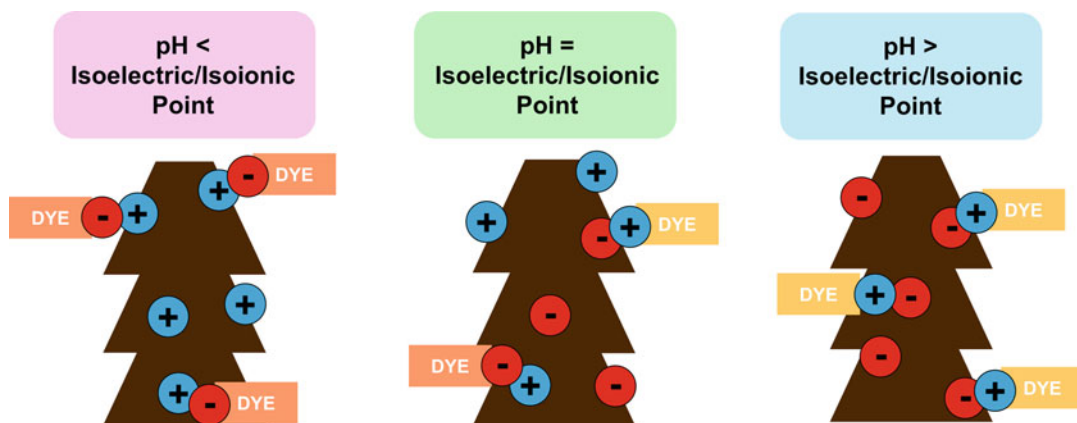
Dye diffusion into hair fibers is largely influenced not only by fiber architecture [24], but also its condition [2, 26] (i.e., if it is damaged/more porous) as well as the molecular size, shape, and polarity of the dye [4], and solvent employed [3]. Solvents can induce hair fiber swelling and thus enhance free volume within the hair fiber and promote the diffusion of dye within. Hair fiber swelling can also be influenced by pH [27]. Swelling is lowest at the isoionic point of hair (5.5 [27]), or the pH at which there is no net protein charge or charge is equilibrated [3, 4, 6, 20, 28]. Contrarily, excess negative or positive protein charge will cause the chains to repel one another (or swelling) [27]. The isoelectric point, or the pH at which the amino acid molecules in hair do not migrate in an electric field (or bear no net electrical charge) [3, 4, 6, 20], can also affect the diffusion of dye into hair. An isoelectric point of 3.67, as determined by Wilkerson and co-workers [29], has mostly been cited in the literature. When probing the dyeability of the surface of hair, the isoelectric point is of more importance, while the isoionic point is more

important when considering the dyeability of the fiber as a whole [20]. At a pH greater than the isoelectric and isoionic points, hair bears a net negative charge [3, 4, 20, 28] and vice versa at a lower pH. Hence cationic dye species display affinity for hair under alkaline conditions [3, 4, 10], and anionic dye species display affinity for hair under acidic conditions (see Fig. 3). Reductive and oxidative chemical treatments and increased free lipid content can decrease the isoelectric point [20], and chemical treatments can also decrease the isoionic point [6, 20]. The diffusion of anionic dyes into hair is slow at pH levels exceeding the isoelectric/isoionic points of hair but faster for nonionic and cationic dyes [2].

Concentration, temperature, and viscosity are other factors that will affect dye diffusion [2]. A higher concentration of dye increases the likelihood of dye interaction with the hair fiber, and the higher the temperature, the faster the diffusion of dye. Too high of a viscosity in the dye formulation will inhibit dye diffusion into the fiber.

Permanent Hair Dye Chemistry

Permanent hair dyes are the most popular of all synthetic dyes. They constitute about 80% of the global hair dye market [30–32] due to their ease of application, longevity, gray hair coverage efficiency (100%), and versatility in colors that can be achieved [3–5, 11]. Their efficacy stems from the diffusion of low molecular weight precursors into hair that, upon oxidation, couple together and

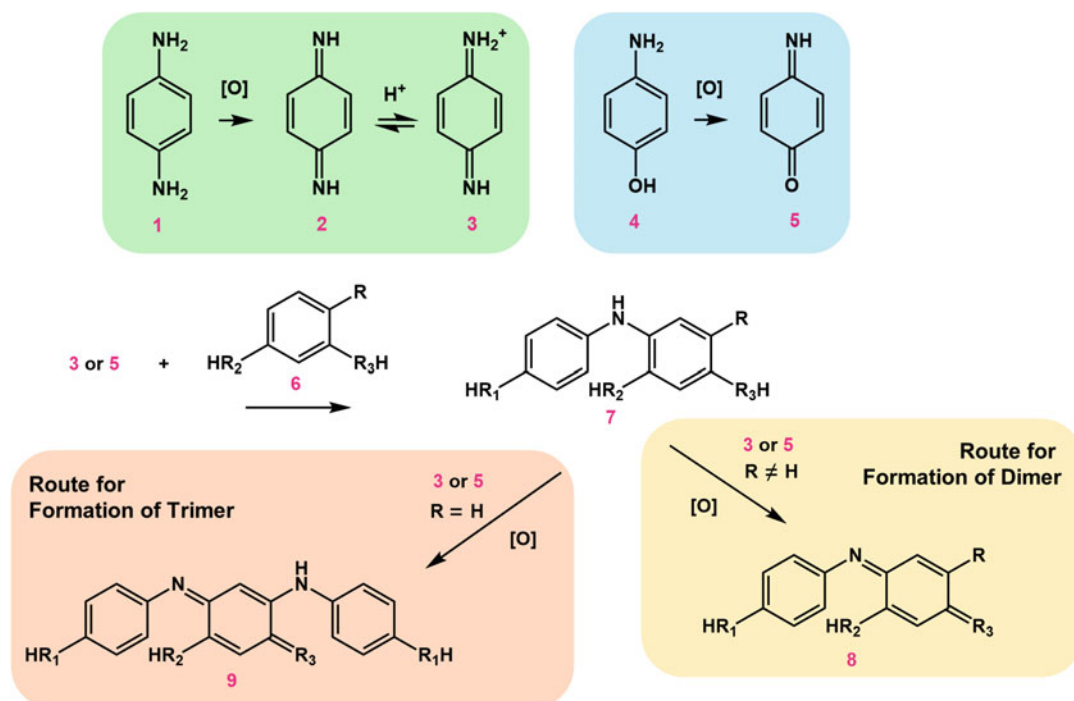


Coloration, Hair, Fig. 3 Schematic representation demonstrating isoelectric and isoionic effects on hair and the affinity of anionic and cationic dyes toward hair at different pH levels

produce colored oligomeric species known as indo dyes [2]. The oligomeric species that form inside the hair become too large to be desorbed (i.e., are mechanically entrapped) [2, 4, 5, 12], promoting wash resistance. Required for permanent hair dye formation are the precursors (primary intermediates and couplers) in addition to oxidizing and alkalizing agents. Primary intermediates are typically *ortho*- or *para*-substituted aromatic hydroxy and amino compounds [1, 3–5], while couplers are typically aromatic compounds possessing electron-rich substituents in the *meta* positions [1, 3]. Some examples of exceptions for couplers are naphthols and pyridines [1]. Examples of other precursors can be found in the Hair Dye Substance Database recently developed by Williams and co-workers [14], which is available for download at https://figshare.com/articles/Hair_Dye_Substance_Database_HDSD/_5505856. Couplers, which are essentially colorless, are known as color modifiers, because they play a role in determining the final color achieved [1, 3, 5, 30]. Direct (non-oxidative) hair dyes may also be included in the formulation of permanent dyes to function as “color agents” that modify the color directly [10], enhancing vibrancy [1]. Examples of color agents are C.I. Disperse Blue 1, 2-nitro-*p*-phenylenediamine (red), and 4-nitro-*o*-phenylenediamine (yellow-orange) [10].

A general reaction scheme for the formation of permanent hair dye is shown in Fig. 4. In summary, *p*-phenylenediamine/PPD (1) or

p-aminophenol/PAP (4) is oxidized using an oxidant such as hydrogen peroxide leading to the formation of *p*-benzoquinonediimine (2) or *p*-benzoquinonemonoimine (5) [5]. The conjugate acid of *p*-benzoquinonediimine (3) or the neutral form of *p*-benzoquinonemonoimine (5) is considered the reactive species that results at alkaline pH [30]. The pK_a of 3 is 5.75 [33]. Therefore, it is the less dominant species in the pH 8–10 range of dyeing, and it decreases in amount by tenfold for each unit increase in pH [33]. An alkalizing agent such as ammonia will activate the oxidizing agent [10] and promote hair swelling [34] and the diffusion of the precursors into the cortex [35]. A subsequent electrophilic substitution reaction with 3 or 5 and coupler (6) will occur at the *para* position to the electron donating group (amino or hydroxyl), resulting in the formation of a leuco (soluble) form of the dinuclear indo dye (7) [3, 5]. If $R \neq H$ in the *para* position, the group present in this position will act as a “blocking group” and inhibit further oligomerization [4, 5, 35], giving rise to the insoluble dinuclear dye species upon further oxidation (8) [36]. If $R = H$, 3 or 5 can undergo 1,4-addition with 7 to form a trinuclear indo dye (9) upon further oxidation [3]. The former case ($R \neq H$, or the formation of dimer) results in light shades, while the latter case ($R = H$, or the formation of trimer) results in dark shades [35]. For examples, *p*-toluenediamine (PTD) and 4-amino-2-hydroxytoluene are



Coloration, Hair, Fig. 4 Oxidative coupling of PPD or PAP with coupler to form permanent hair dye. (Note: R₁, R₂, and R₃ are O or NH)

expected to form dimers, while PTD and *m*-aminophenol are expected to form trimers [37]. Isolated dyes from hair have also been of larger molecular weight species [10].

As many as 10–12 precursors can exist within permanent hair dye formulations [3]. Generally, dyes affording darker colors contain the highest concentration of precursors, while light colors contain the lowest concentrations [1]. Initial concentration of dye to form the target colors varies from 0.006 to 7%. The final concentration of dye is diluted upon the use of a developer (contains oxidant), for example, by 50% (for 1:1 mixing). Multiple, competitive reactions occur when there are more than one primary intermediate-coupler pair in the formulation. As a result, the final color seen depends a variety of factors such as the reactivity and concentration of the couplers [33], the diffusion rates and redox potential of the precursors [4], the pH at which dyeing occurs [33], and the starting base color of the hair. Consequently, the practice of hair coloration can be an art.

Color formation is more rapid in the hair fiber than in solution, because peroxide in aqueous solution is a weak oxidant (HOO^-) for primary intermediates like PPD and PAP, and species that are more reactive can form outside of solution [30, 33]. Since the reactions are quite fast, the indo dyes form primarily in the cuticle and outer portion of the cortex, because the precursors do not have time to diffuse deep into the core of the fiber before reacting [10]. The dye, though mechanically entrapped within the hair fiber, can bond to hair through secondary and ionic bonds [2]. Covalent bonding of dye is also possible between nucleophilic regions in hair [10].

Permanent Hair Dye Toxicity Concerns

Permanent hair dye precursors have raised potential toxicity concerns for many years. Indeed, their carcinogenic potential has been a matter of debate since the 1975 Ames study that found a large

number of hair dye products to contain substances that were mutagenic [38]. As a result, there was a major reformulation period during the 1970s and 1980s that resulted in the removal of substances from hair dye products demonstrating mutagenicity [36]. In 2010, the International Agency for Research on Cancer (IARC) published a monograph detailing an extensive epidemiological data review of hair dyes and cancer risk in animals and humans [1]. It was concluded that there is limited evidence from studies involving laboratory animals and occupational workers and not enough evidence in consumers to establish the unequivocal carcinogenic risk of hair dyes. Thus, hair dyes have been designated as “probably carcinogenic to humans” (Group 2A) for occupational workers and “not classifiable to its carcinogenicity to humans” (Group 3) for consumers. The Food and Drug Administration (FDA) also agrees that there is not enough evidence to conclude that hair dyes are carcinogenic to consumers [11, 39]. In addition, the Scientific Committee on Consumer Safety (SCCS) acknowledges that there is no clear evidence of an elevated cancer risk for the use of dyes marketed in the European Union [11].

On the other hand, several precursors are strong/extreme skin sensitizers (or allergens) [40]. Indeed, a recent United States consumer exposure study conducted by Hamann and co-workers identified 106 out of 107 (99%) of commercial hair dye products contained at least one allergen, the average product containing six allergens [41]. Allergens found in most of the products included resorcinol (in 89% of the products), *p*-phenylenediamine/PPD (in 78% of the products), *m*-aminophenol/MAP (in 75% of the products), and *p*-aminophenol/PAP (in 60% of the products). PPD and other precursors can cross-react with benzocaine, black rubber, and other frequently encountered chemicals [41, 42], thereby enhancing skin sensitization potential [43]. Bleaching agents, sulfites, fragrances, and preservatives found in the product formulations can also induce consumer sensitivities [42]. Symptoms such as swelling, inflammation, itching, asthma, renal failure, headache, insomnia, and dizziness [34] can occur within 2–48 h or more, depending on the severity of exposure [41,

42]. Common areas affected include the forehead, neck, and ears. Generally, the scalp itself is not affected, although dye is sometimes located here [42]. Some consumers are more prone than others to develop allergies. For example, likelihood increases with age in addition to frequency of use. In regions where the dyeing of beard hair is more common, men are more likely than women to develop allergic responses, but typically women are more prone to developing allergies since they use hair dyes more frequently. Consumers that color their hair using darker shades are also more likely to develop an allergy, because the associated products contain the highest concentrations of allergens [43].

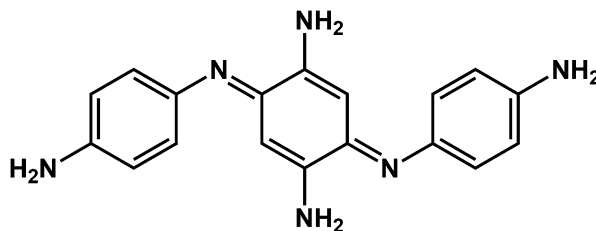
Unfortunately, there is no “cure” to hair dye allergies, although temporary relief of symptoms can be obtained through the use of medications [42]. The consumer is simply advised to avoid dyes that they may have sensitivities toward or to avoid the use of permanent hair dyes altogether [44]. However, there are certain challenges with heeding this advice. For example, the North American baseline patch test series only includes PPD, but other hair dye product components may induce allergenic reactions [41]. Consequently, hair dye allergies may be undiagnosed. Also, the consumer may be confused by terms that may be used to market hair dye products, such as “natural,” “healthy,” and “protect”. Some of these products could still cause skin sensitivity.

The environmental effects of hair dyes cannot be ignored, since their manufacture and consumption generates a large amount of waste [45]. Indeed, PPD, *p*-toluenediamine (PTD), resorcinol, and Bandrowski’s base (see Fig. 5), a known mutagen [38] that arises from the self-coupling of PPD, was found in wastewater collected from a hair dyeing facility [46, 47].

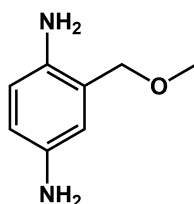
Alternative Permanent Hair Dye Precursors and Coloration Technologies

Alternative permanent hair dye precursors and coloration technologies have been proposed to eliminate permanent hair dye safety concerns, many of which have been extensively reviewed

Coloration, Hair,
Fig. 5 Structure of
 Bandrowski's base



Coloration, Hair,
Fig. 6 Chemical structure
 of 2-methoxymethyl-PPD



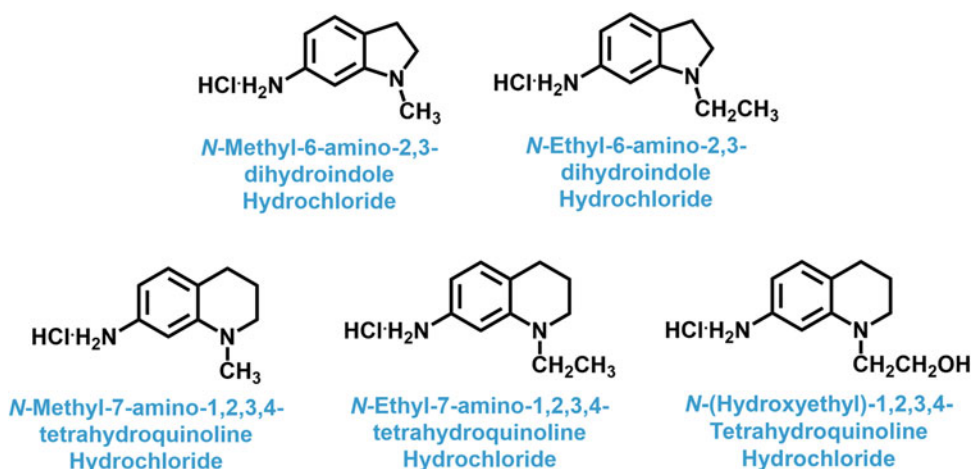
by Morel and Christie [3]. Alternative permanent hair dye precursors include substituted derivatives of PPD and PAP, mono-, di-alkyl and 6-substituted resorcinols, heterocyclic-based pyrimidines and pyridines, and pyrazoles. Benzene ring substitution in the *ortho*-position of *p*-phenylenediamines and *p*-aminophenols was one strategy attempted to reduce the toxicity of these compounds while enhancing their color performance on hair. Some new strategies to oxidize precursors without the need for harsh oxidizing agents were also reviewed. Few newly proposed alternative precursors have had commercial success. One example is 2-methoxymethyl-PPD [11], which introduces a methoxymethyl side chain onto PPD [48] (see Fig. 6). This alternative primary intermediate was recently commercialized due to its excellent color performance on hair and moderate sensitization potential compared to PPD and PTD [48], even for some individuals demonstrating PPD and PTD allergies [49].

Li and co-workers synthesized a series of tetrahydroquinoline and dihydroindole couplers (see Fig. 7) reacted with either PPD, PTD, or PAP to afford blue or red shades [35]. The tetrahydroquinoline derivatives, in particular, *N*-methyl-7-amino-1,2,3,4-tetrahydroquinoline, exhibited color vibrancy and superior wash resistance and stability to light compared to the dihydroindole derivatives and commercial hair

dyes studied. The toxicological profile of the new couplers was not reported, though.

Alternative permanent coloration technologies, despite some individual successes, have yet to prevail over the economy and efficacy of conventional permanent hair dye precursors. One interesting proposed alternative permanent hair coloration technology involves the use of infused biocompatible spherical silica nanoparticles (206 nm average diameter) with commercial red acid dye (Telon Red M-BL) [26]. Sampaio and co-workers adopted this technology and found the nanoparticles penetrated only bleached hair fibers. Coloration was performed at pH 4.5, which afforded more potential protein binding sites for the colorant. Significant color loss was seen during the first six washes, due to dye removal from the outermost layers of the hair fiber. Beyond six washes, color loss was minimal up to thirteen washes.

Hair coloration strategies have been developed to take advantage of the polymerization of dopamine on a solid surface with the aid of a metal-complexing agent. This technology mimics the natural hair coloration formation process (see **Melanin** chapter). One advantage of this technology is the low skin sensitization potential of polydopamine (PDA), which has been found to cause no edema or erythema in the rabbit skin protocol [50]. Another benefit is the strong binding interaction potential of complexed PDA with hair keratin, thereby enhancing the wash resistance of the colorant. In a study conducted by Im and co-workers, a PDA metal complex was developed on hair with the aid of FeSO₄ [51]. The resultant complex was black in color, and little color change was observed for colored hair fibers that were soaked in a soap solution up to 30 min. Other colors on hair could be achieved by replacing the



Coloration, Hair, Fig. 7 Alternative permanent hair dye precursors developed by Li and co-workers [35]

Fe^{2+} with an ion such as Cu^{2+} . More recently, Gao and co-workers developed a PDA metal complex on hair with the aid of CuSO_4 and H_2O_2 [50]. The resultant complex was black in color, rapidly formed (ca. 5 min), and was wash resistant up to 30 shampoos.

Another interesting hair dyeing technology was proposed by Ochiai and co-workers, who examined the viability of certain acid dyes as novel permanent hair colorants when applied to hair with metal salts such as AlCl_3 [13]. Acid dyes studied included C.I. Acid Orange 7 and C.I. Acid Red 52, which are traditionally non-oxidative hair dyes. Reportedly, the dyes outperformed conventional semipermanent and permanent dyes in terms of color brightness and longevity. The toxicological profile of the dyes was not reported, though.

Some other interesting potential permanent hair coloration technologies include the use of graphene-based dye [52], the use of catechinone, a water-soluble colorant obtained from tyrosinase reaction with (+)-catechin [45], and the addition of coloring agents to the hair follicle during growth and development [5]. The use of photonics to develop color on hair without the presence of a physical colorant is a research avenue yet to be explored. Although successful research efforts to inhibit or reverse the process of graying may decrease the need for the design of alternative

hair dyes, such endeavors have proven quite challenging to achieve [11]. Despite the above exciting developments, there remains a need for alternative permanent hair coloration strategies that are environmentally friendly. Thus, studies in this important arena continue. This work includes the use of cheminformatics modeling to facilitate the screening of large dye databases for potential hair dye alternatives [53].

Cross-Reference

- [Colorant, Environmental Aspects](#)
- [Colorant, Natural](#)
- [Dye](#)
- [Melanin](#)

References

1. Some aromatic amines, organic dyes, and related exposures: occupational exposures of hair dressers and barbers and personal use of hair colourants. IARC Monogr. Eval. Carcinog. Risks Hum. **99**, 499–658 (2010)
2. Underwood, D.L.: Basic elements of dyeing human hair. J. Soc. Cosmet. Chem. **12**, 155–162 (1961)
3. Morel, O.J.X., Christie, R.M.: Current trends in the chemistry of permanent hair dyeing. Chem. Rev. **111**, 2537–2561 (2011)

4. Christie, R.M., Morel, O.J.X.: The coloration of human hair. In: Lewis, D.M., Rippon, J.A. (eds.) *The Coloration of Wool and Other Keratin Fibers*, pp. 357–391. Wiley, Chichester (2013)
5. Christie, R.M.: Colour in cosmetics, with special emphasis on hair coloration. In: *Colour Chemistry*, pp. 250–265. Royal Society of Chemistry, Cambridge (2015)
6. Morel, O., Christie, R., Greaves, A., Morgan, K.: Enhanced model for the diffusivity of a dye molecule into human hair fibre based on molecular modelling techniques. *Color Technol.* **124**, 301–309 (2008)
7. Brown, K.C., Marlowe, E., Prota, G., Wenke, G.: A novel natural-based hair coloring process. *J. Soc. Cosmet. Chem.* **48**, 133–140 (1997)
8. Greaves E., Greaves, J.T.: Composition for dyeing keratin fibers and a method of dyeing hair using same. US Patent US20070251024A1. (2007)
9. Harrison, S., Sinclair, R.: Hair colouring, permanent styling and hair structure. *J. Cosmet. Dermatol.* **2**, 180–185 (2003)
10. Robbins, C.R.: Dyeing Human Hair. In: *Chemical and Physical Behavior of Human Hair*, pp. 445–488. Springer, Berlin/Heidelberg (2012)
11. Rust, R.C., Schlatter, H.: Hair Dyes. In: Draelos, Z. (ed.) *Cosmetic Dermatology: Products and Procedures*, pp. 239–249. Wiley, Chichester (2016)
12. Zhang, G., McMullen, R.L., Kulcsar, L.: Investigation of hair dye deposition, hair color loss, and hair damage during multiple oxidative dyeing and shampooing cycles. *J. Cosmet. Sci.* **67**, 1–11 (2016)
13. Ochiai, M., Kawasoe, T., Yasuda, M., Harada, Y., Kimura, T., Kambe, T., Kurita, J.: A novel "permanent" acid-type hair color made possible with dye-metal ion complex technology. *J. Cosmet. Sci.* **56**, 29–46 (2005)
14. Williams, T.N., Kuenemann, M.A., Van Den Driessche, G.A., Williams, A.J., Fourches, D., Freeman, H.S.: Toward the rational design of sustainable hair dyes using cheminformatics approaches: step 1. Database development and analysis. *ACS Sustain. Chem. Eng.* **6**, 2344–2352 (2018)
15. Leeder, J.D.: Comments on "pathways for aqueous diffusion in keratin fibers". *Text. Res. J.* **69**, 229 (1999)
16. Rippon, J.A.: Comments on "pathways for dye diffusion in wool fibers". *Text. Res. J.* **69**, 307–308 (1999)
17. Swift, J.A.: Further comments on "pathways for aqueous diffusion in keratin fibers". *Text. Res. J.* **70**, 277–278 (2000)
18. Swift, J.A.: Pathways for aqueous diffusion in keratin fibers. *Text. Res. J.* **69**, 152 (1999)
19. Wortmann, F., Wortmann, G., Zahn, H.: Pathways for dye diffusion in wool fibers. *Text. Res. J.* **67**, 720–724 (1997)
20. Robbins, C.R.: Interactions of shampoo and conditioner ingredients with hair. In: *Chemical and Physical Behavior of Human Hair*, pp. 329–444. Springer, Berlin/Heidelberg (2012)
21. Gummer, C.L.: Elucidating penetration pathways into the hair fiber using novel microscopic techniques. *J. Cosmet. Sci.* **52**, 265–280 (2001)
22. Sideris, V., Holt, L.A., Leaver, I.H.: A microscopical study of the pathway for diffusion of rhodamine B and octadecylrhodamine B into wool fibres. *J. Soc. Dyers Color.* **106**, 131–135 (1990)
23. Leeder, J.D., Rippon, J.A., Rothery, F.E., Stapleton, I. W.: Use of the transmission electron microscope to study dyeing and diffusion processes. *Proc. 7th Int. Wool Text. Res. Conf.* **5**, 99–108 (1985)
24. Leeder, J.D., Holt, L.A., Rippon, J.A., Stapleton, I.W.: Diffusion of dyes and other reagents into the wool fibre. *Proc. 8th Int. Wool Text. Res. Conf.* **4**, 227–238 (1990)
25. Kelch, A., Wessel, S., Will, T., Hintze, U., Wepf, R., Wiesendanger, R.: Penetration pathways of fluorescent dyes in human hair fibres investigated by scanning near-field optical microscopy. *J. Microsc.* **200**, 179–186 (2000)
26. Sampaio, S., Maia, F., Gomes, J.: Diffusion of coloured silica nanoparticles into human hair. *Color Technol.* **127**, 55–61 (2011)
27. Popescu, C., Höcker, H.: Hair—the most sophisticated biological composite material. *Chem. Soc. Rev.* **36**, 1282–1291 (2007)
28. Carey, F.A.: Amino acids, peptides, and proteins. In: *Organic Chemistry*, pp. 1106–1161. McGraw-Hill, New York (2008)
29. Wilkerson, V.A.: The chemistry of human epidermis II. The isoelectric points of the stratum corneum, hair, and nails as determined by electrophoresis. *J. Biol. Chem.* **112**, 329–335 (1935)
30. Corbett, J.F.: 10 synthetic dyes for human hair. In: Freeman, H.S., Peters, A.T. (eds.) *Colorants for Non-Textile Applications*, pp. 456–477. Elsevier, Amsterdam (2000)
31. Antunes, E., Cruz, C.F., Azoia, N.G., Cavaco-Paulo, A.: The effects of solvent composition on the affinity of a peptide towards hair keratin: experimental and molecular dynamics data. *RSC Adv.* **5**, 12365–12371 (2015)
32. Robbins, C.R.: Chemical composition of different hair types. In: *Chemical and Physical Behavior of Human Hair*, pp. 105–176. Springer, Berlin/Heidelberg (2012)
33. Corbett, J.: Chemistry of hair colorant processes - Science as an aid to formulation and development. *J. Soc. Cosmet. Chem.* **35**, 297–310 (1984)
34. Kim, K., Kabir, E., Jahan, S.A.: The use of personal hair dye and its implications for human health. *Environ. Int.* **89–90**, 222–227 (2016)
35. Li, D., Huang, Y., Su, J.: Dyeing behaviours of amino heterocyclic compounds as blue oxidative hair dye precursors applied to keratin fibres. *Int. J. Cosmet. Sci.* **33**, 183–189 (2011)
36. Corbett, J.F.: An historical review of the use of dye precursors in the formulation of commercial oxidation hair dyes. *Dyes Pigments.* **41**, 127–136 (1999)

37. SCCS: Opinion on reaction products of oxidative hair dye ingredients formed during hair dyeing processes, SCCS/1311/10, pp. 1–76 (2010)
38. Ames, B.N., Kammen, H.O., Yamasaki, E.: Hair dyes are mutagenic: identification of a variety of mutagenic ingredients. *Proc. Natl. Acad. Sci. USA*. **72**, 2423–2427 (1975)
39. Hair Dyes: US FDA. <http://www.fda.gov/Cosmetics/ProductsIngredients/Products/ucm143066.htm>. Accessed 19 Aug 2019
40. Boonchai, W., Bunyavaree, M., Winayanuwattikun, W., Kasemsarn, P.: Contact sensitizers in commercial hair dye products sold in Thailand. *Contact Derm.* **74**, 222–229 (2016)
41. Hamann, D., Hamann, C.R., Thyssen, J.P., Lidén, C.: p-Phenylenediamine and other allergens in hair dye products in the United States: a consumer exposure study. *Contact Derm.* **70**, 213–218 (2014)
42. McFadden, J., Dyes, H.: In: Johansen, J.D., Lepoittevin, J., Thyssen, J.P. (eds.) *Quick Guide to Contact Dermatitis*, pp. 189–193. Springer, Berlin Heidelberg (2016)
43. Gray, J.: The safety of hair dyes. In: *The World of Hair Colour*, pp. 111–114. Thomas Learning, London (2005)
44. Pongpairroj, K., McFadden, J.P., Basketter, D.A.L.: Advice for patients with hair dye allergy remains ‘stop using permanent hair dyes’. *Br. J. Dermatol.* **174**, 957–958 (2016)
45. Hidekazu, Y., Akiko, T., Kazue, I., Masahisa, U., Hiroshi, U.: Hair dyeing by using catechinone obtained from (+)-catechin. *J. Cosmet. Dermatol. Sci. Appl.* **2**, 158–163 (2012)
46. Bessegato, G.G., De Souza, J.C., Cardoso, J.C., Zanoni, M.V.B.: Assessment of several advanced oxidation processes applied in the treatment of environmental concern constituents from a real hair dye wastewater. *J. Environ. Chem. Eng.* **6**, 2794–2802 (2018)
47. de Souza, J.C., da Silva, B.F., Morales, D.A., de Aragão Umbuzeiro, G., Zanoni, M.V.B.: Assessment of the autoxidation mechanism of p-toluenediamine by air and hydrogen peroxide and determination of mutagenic environmental contaminant in beauty salon effluent. *Sci. Total Environ.* **685**, 911–922 (2019)
48. Goebel, C., Troutman, J., Hennen, J., Rothe, H., Schlatter, H., Gerberick, G.F., Blömeke, B.: Introduction of a methoxymethyl side chain into p-phenylenediamine attenuates its sensitizing potency and reduces the risk of allergy induction. *Toxicol. Appl. Pharmacol.* **274**, 480–487 (2014)
49. Kock, M., Coenraads, P., Blömeke, B., Goebel, C.: Continuous usage of a hair dye product containing 2-methoxymethyl-para-phenylenediamine by hair-dye-allergic individuals. *Br. J. Dermatol.* **174**, 1042–1050 (2016)
50. Gao, Z.F., Wang, X.Y., Gao, J.B., Xia, F.: Rapid preparation of polydopamine coating as a multifunctional hair dye. *RSC Adv.* **9**, 20492–20496 (2019)
51. Im, K., Kim, T., Jeon, J.: Metal-chelation-assisted deposition of polydopamine on human hair: a ready-to-use eumelanin-based hair dyeing methodology. *ACS Biomater. Sci. Eng.* **3**, 628–636 (2017)
52. Luo, C., Zhou, L., Chiou, K., Huang, J.: Multifunctional graphene hair dye. *Chem.* **4**, 784–794 (2018)
53. Williams, T.N., Van Den Driessche, G.A., Valery, A.R. B., Fourches, D., Freeman, H.S.: Toward the rational design of sustainable hair dyes using cheminformatics approaches: step 2. Identification of hair dye substance database analogs in the max weaver dye library. *ACS Sustain. Chem. Eng.* **6**, 14248–14256 (2018)

Coloration, Ink-Jet Printing

Yi Ding¹ and Lisa Chapman²

¹Donghua University, Shanghai, China

²North Carolina State University, Raleigh, NC, USA

Synonyms

Continuous Ink Jet (CIJ); Cyan, Magenta, Yellow, or Black (CMYK); Direct to garment (DTG); Drop-on-Demand (DOD); Piezoelectric ink-jet (PIJ); Raster image processing (RIP); The international textile machinery exhibition (ITMA); Thermal ink-jet (TIJ)

Definition

Textile ink-jet printing is a noncontact method of jetting colorants directly onto fabric. This technology can be used for roll-to-roll, or direct-to-garment (DTG) printing, and is a rapidly emerging alternative to screen printing, particularly for sampling and smaller runs. By eliminating the screen preparation process, ink-jet printing allows more flexibility in the printing process, provides capability for fine line detail, a near infinite number of colors, freedom from repeats, and the ability to engineer a print design within a garment or furniture marker. The fine line detail and increased number of colors can produce a near photographic

image on the textile substrate, enabling new product categories.

Ink-jet machines are typically classified by printhead and are continuous ink-jet (CIJ) and drop-on-demand ink-jet (DOD). With a CIJ printer, the ink is projected through the nozzles on the printhead at a constant speed, and under a constant pressure. With a DOD printer, the ink drops are ejected only when needed. The two main mechanisms used to generate drops in a DOD system are piezoelectric ink-jet (PIJ) and thermal ink-jet (TIJ).

Technology Development

Ink-jet technology was first described by Lord Rayleigh in 1878 [1], as a “nonimpact dot matrix printing technology in which droplets of ink are jetted from a small aperture directly to a specific position on a medium to create an image” [1, 2]. Commercialization of this technology was first seen in paper printing applications and drove many of the advances in ink-jet printing technologies throughout much of the mid-twentieth century. In the early 1970s, the first printers were developed specifically for the textiles industry; most notably Milliken Company introduced the Millitron printer for carpet printing. The Millitron, had fairly low print resolution, only 20 dots per inch (dpi), however this resolution was suitable for pile fabrics [3]. In the mid/late twentieth century, systems for digital ink-jet printing for textile substrates were developed, with sampling as the main use [4].

As demand for textile ink-jet printing increased, advancements occurred in print resolution and printing speed. The ink-jet printer first

used for roll-to-roll textiles was based on dye jets controlled by solenoid valves. Introduced in 1987, this printer had CIJ heads, and operated on the binary principle with four printheads (cyan, magenta, yellow, and black) that scanned across the fabric attached to a rotating drum. Print production average done hour per square meter. Multilevel charge systems were introduced later on with a multiple array of printheads to increase output [5]. In 1999, a TIJ water-based ink printer was launched with a 300 dpi resolution, and the print speed could reach 12 m²/h. In 2003, the Artistic 2020 printer premiered at the ITMA expo. This printer was suitable for reactive, disperse, acid dyes and pigment inks, with 600 dpi capability, and a print speed of 30 m²/h [6, 7].

Currently, ink-jet printers can be classified into four categories, based on speeds ranging from 4 m²/h to greater than 120 linear m/min (see Table 1). Digital textile printers that employ single-pass printhead systems print on the substrate as the fabric is fed under a fixed print unit. Single-pass printers print at significantly higher speeds (up to 120 linear m/min) than scan-type printer that print by moving the printhead carriage back and forth. In single-pass printing, each printhead station jets one color, cyan, magenta, yellow, or black (CMYK), and up to four additional stations of colors can be added such as blue, green, violet, grey, red, or orange.

The footprint of single-pass printers is considerably larger than scanning mode printers; however, the speed of single-pass printers justifies the large square footage, as these printers are capable of printing up to 70 m²/min. By ITMA 2019, single-pass printers achieved speeds of 120 linear m/min and had resolutions of up to 1200 dpi and

Coloration, Ink-Jet Printing, Table 1 Printer classifications by production speed

Printer	Speed	Colorant capability				
		Acid	Disperse	Pigment	Reactive	Vat
Class 1	4 m ² /h to 12.5 m ² /h	x	x	x	x	x
Class 2	16.5 m ² /h to 33.5 m ² /h	x	x	x	x	x
Class 3	58.5 m ² /h to 334.5 m ² /h	x	x	x	x	*
Class 4	70 m ² /min	x	x	x	x	*

*Under development



Coloration, Ink-Jet Printing, Fig. 1 Double-sided digitally printed silk fabric from Atexco

some even had capability of double-sided printing [8], see Fig. 1.

System Components

There are several components that comprise a textile ink-jet printing system:

1. An assembly of inkjet printheads, which generates streams of microscopic ink droplets and directs them to their target position on the fabric. The ink can be pressurized at 3–5 MPa for the CIJ printhead but is supplied at essentially atmospheric pressure and is most commonly used for DOD printheads [5]. The majority of piezo-electric printheads used in textile ink-jet printers are manufactured by companies such as Seiko Epson, Ricoh, Konica Minolta, Kyocera, and Fuji Dimati [8]. HP manufactures thermal printheads for DOD printers.
2. A mechanism that feeds and presents the fabric to the traversing ink-jet printheads ensures perfect registration and alignment throughout the fabric, even for delicate and unstable substrates such as knits or fine silks. The printing system may also have functions to preheat, dry, and/or heat-set the printed fabric after printing, with

capability of rolling up the output fabric smoothly with even tension.

For roll-to-roll printing, the fabric roll is unwound at the same speed as the printer and wound under tension to another roll after printing. With difficult-to-handle fabrics, for example, lightweight fabrics such as 5–10 mm silks, a stiff paper backing is lightly adhered to the pretreated fabric, to facilitate the fabric feed in the roll-to-roll printing system. More recently, a sticky-belt fabric feeding system was developed, which utilizes a special chemical-coated belt to support fabrics, reducing distortion or shrinkage of the substrate [9].

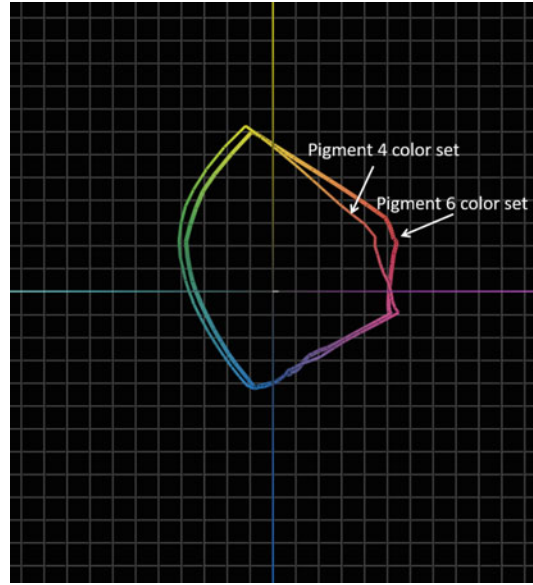
3. Software, including printer drivers, raster image processing (RIP), and color management systems, is needed to convert computer-based designs into the electronic signals which control the ink-jet printhead and machine. The software also has color management functions to ensure reproducible results with different batches of fabric, and to provide a total interface with the other components of a digital design, sampling, and production environment.
4. Ink-jet colorants, whether pigments or dyes, need to be milled and filtered to much finer tolerances than colorant used for conventional screen or roller printing. The inks for textile ink-jet printing must exhibit key physical and chemical properties to be suitable for jetting, including purity, viscosity, surface tension, conductivity, pH values, foaming properties and particle size, temperature response characteristics, and long shelf life without settling or mold growth [10]. Inks must not corrode surfaces or form deposits within the printhead. Also, the chemical properties of the colorant determine the extent of the interaction with the textile substrate, which is enhanced by pre- and post-treatment steps [11–13], see Fig. 2. Specifically, the ink must be readily and evenly adsorbed onto the individual fibers and yarns of a fabric after printing, providing the required color and intensity characteristics. Properties such as washing, light, and rubbing fastness and wicking resistance are also enhanced.



Coloration, Ink-Jet Printing, Fig. 2 Digitally printed polyester fabric before (top) and after (bottom) pretreatment

Commercially available colorants in textile ink-jet inks include reactive, acid, and disperse dyes, nano-scaled pigments, and vat dyes. A four-color set comprised of cyan, magenta, yellow, and black (CMYK) has become the industry standard for textile ink-jet printing. Additional colors can be added to the standard ink set, to increase the color gamut and reduce total ink consumption [9, 14]. An example of the color gamut derived from a four- and six-color set is shown in Fig. 3.

5. Textile substrates employed for printing normally require a pretreatment step because the fabric construction (weave or knit) can cause wicking problems, leading to a higher ink penetration and reduced color strength on the substrate [5]. Pretreatment and a subsequent post-treatment step improve print quality [9], providing the adhesion and reactivity needed to meet end-use specifications such as target color and fastness properties.



Coloration, Ink-Jet Printing, Fig. 3 Color gamut comparison for representative four- versus six-pigment color sets

Textile ink-jet printing offers opportunities for exciting new products to enter a variety of apparel, garment badges and labels, interior textiles, carpets and floor coverings, technical textiles, flags and banners, and industrial graphics markets [15]. This technology also offers opportunities to capture personalization and quick response to fashion changes, as it provides lower costs for shorter print runs and just-in-time delivery [16].

The development of high-speed digital (ink-jet) printers provides a process that is competitive with rotary screen printing in terms of speed and cost, while offering a competitive advantage when printing shorter runs or multiple color ways [17]. Also, no screens are required, which means all costs to engrave screen, make paste, strike-offs, downtime, and wastage are eliminated. Further, less space is required for industrial printing, with no need for operations associated with film masters, stencils, screens, or plates. The multicolor registration of ink makes possible an unlimited repeat size and number. Inventory and pollution control costs are decreased tremendously. Therefore, ink-jet

printed fabrics would offer a game changing strategy for the international textile printing industry.

Cross-References

- [Coloration, Ink-Jet Printing](#)
- [Coloration, Textile](#)

References

1. Rayleigh, L.: On the instability of jets. *Proc. Lond. Math. Soc.* **1**(1), 4–13 (1878)
2. Sirringhaus, H., Shimoda, T.: Inkjet printing of functional materials. *MRS Bull.* **28**(11), 802–806 (2003)
3. Tincher, W.C., Hu, Q., Li, X.: Ink jet systems for printing fabric. *Textile Chemist Colorist.* **30**(5), 24–27 (1998)
4. Tyler, D.J.: Textile digital printing technologies. *Text. Prog.* **37**(4), 1–65 (2005)
5. Ujiie, H. (ed.): *Digital printing of textiles*. Woodhead, Cambridge (2006)
6. Heinzl, J., Hertz, C.: Ink-jet printing. *Adv. Electron. Electron Phys.* **65**, 91–171 (1985)
7. Moser, L.S.: ITMA 2003 review: textile printing. *J. Textile Apparel Technol. Manage.* **3**(3), 1–15 (2003)
8. Provost, J.: Developments in digital textile pigment inks for high-productivity machines. *Digital Textiles.* **4**, 12–14 (2015)
9. Cie, C.: *Ink jet textile printing*. Woodhead, Cambridge (2015)
10. Fu, S., et al.: Development of inkjet printing ink for textiles. *China Textile Leader.* **4**(4), 34–37 (2012)
11. Fu, S., et al.: Preparation of encapsulated disperse dye dispersion for polyester inkjet printing ink. *J. Appl. Polym. Sci.* **121**(3), 1616–1621 (2011)
12. Chao, Y., Chang, S.: Preparation of microemulsion based disperse dye inks for thermal bubble ink jet printing. *J. Imag. Sci. Technol.* **51**(5), 413–418 (2007)
13. Freeman, H.S., Posey Jr., J.C.: An approach to the design of lightfast disperse dyes-analogs of disperse yellow 42. *Dyes Pigments.* **20**(3), 171–195 (1992)
14. Ding, Y., Freeman, H.S., Parrillo-Chapman, L.: Color gamut analysis and color fastness evaluation for textile inkjet printing application on polyester. *J. Imag. Sci. Technol.* **61**(5), 50503-1–50503-8 (2017)
15. Dawson, T.: Ink jet printing of textiles: an overview of its development and the principles behind ink drop formation and deposition. *Colourage.* **51**(10), 75–82 (2004)
16. Gupta, S.: Inkjet printing-a revolutionary ecofriendly technique for textile printing. *Indian J. Fibre Textile Res.* **26**(1–2), 156–161 (2001)
17. Dawson, T., Hawkyard, C.: A new millennium of textile printing. *Color. Technol.* **30**(1), 7–20 (2000)

Coloration, Mordants

Martin Bide

Department of Textiles, University of Rhode Island, Fashion Merchandising and Design, Kingston, RI, USA

Definition

Probably the one word that those otherwise unfamiliar with dyeing know is “mordant,” and it is derived from the Latin word “mordere” (to bite) with the implication that such a compound forms a link between dye and fiber. However, given the complexities of the dyeing process, it is difficult to write a simple definition for “mordant.” The concept is rooted in history, when only natural fibers and natural dyes were available, and scientific understanding was insufficient to explain fully the functions of the various materials used to provide dyeings that were fast to washing, sunlight, etc. As that understanding has developed, the need for mordants has declined to the point where they are commercial oddities. For the purposes of this chapter, the definition is as follows:

A mordant is a substance applied to a textile substrate in parallel with the dyeing process that modifies the interaction of dye and fiber (and remains present in the subsequent dyed material) to provide better uptake, better fastness, and/or a wider range of colors than would otherwise be achieved.

Nevertheless, there are substances that fall into this definition which are not called mordants and some substances described as mordants that do not fit the definition. O'Neill in the 1860s [1] suggested that a mordant should “exert an affinity for the fibrous material to which it is applied, and,

... an attraction for coloring matters,” and this is largely in agreement with the definition given here but implies that the attractions are simultaneous which is rarely the case. Shore [2] gives a definition based on the twentieth-century commercial practice but recognizes the implicit ambiguities.

From the above definition, it is clear that mordants compensate for some lack of substantivity or fastness of a dye. The watershed between a

dyeing world that relied heavily on mordants and one that was becoming relatively mordant-free coincides with the move from natural to synthetic dyes. Mordants were the means of getting good results from natural dyes that were not “designed” for dyeing textiles. Mordants were used for centuries to achieve remarkable results. That strong historical aspect complicates the definition, as our more recent understanding of the chemistry behind the dyeing process has revealed that substances formerly included in the broad scheme of mordants do not truly behave as such.

For several decades prior to the full development of synthetic dyes, chemical knowledge was increasing, and as new elements were identified and new compounds prepared, a wide range of substances were tested as mordants to expand the coloration technology of natural dyes. Even after the first synthetic dyes were developed, their usefulness was broadened by the use of mordanting materials. A notable development occurred in the 1880s, when a new group of synthetic dyes was named “substantive dyes” (later “direct dyes”), because they were substantive toward and dyed cotton “directly” *without* the use of a mordant. Since then, the skill of the chemist has obviated the need for mordants by creating dyes that not only have color but also have the molecular features required to provide dye-fiber substantivity, ultimate insolubility, or functional groups to react covalently with the fiber. The result is a much-simplified (and more efficient) dyeing process that readily provides fast colors.

The (synthetic) dyes form groups that were codified in the second edition of the Colour Index in 1956 [3]. One such group was “mordant dyes.” These dyes were applied to wool in conjunction with a chromium mordant, but their use has lessened as environmental concerns over the use of heavy metals, such as chromium, increase. Together with the very limited use of natural dyes, it is not surprising that mordants are a rarity in (commercial) dyeing today. Current use of mordants is largely confined to home and craft dyeing.

Synthetic dyes were well established before the development of synthetic fibers, and thus when

such fibers were introduced, dyers had mordant-free processes with which to dye them. With one or two exceptions, mordants have been almost exclusively associated with the dyeing of natural fibers. In some cases when materials have been used to improve the interaction of dyes and synthetic fibers, they have not been referred to as mordants, even though they fulfilled essentially the same function.

Overview

In a dyeing process, dyes are soluble (or sparingly soluble), and a (textile) substrate (fiber, yarn, fabric) absorbs the dye, through forces of attraction, from an external solution [4] (► [Coloration, Textile](#)). The sum of the attracting forces is referred to as substantivity. The concept of “substantivity,” relating to dyes that impart color without the use of a mordant, is attributed to Bancroft [5] and significantly predates the development of synthetic dyes, even though inherent substantivity is rare in natural dyes. Substantivity also contributes to the subsequent resistance of the dye to removal under in-use conditions (► [Coloration, Fastness](#)).

A dyer seeks to achieve a steady and even uptake of dye by the substrate. For a given dye/substrate/machine, this is achieved with control of temperature, agitation, and liquid volume, augmented with dyebath additives. Such adjuncts to the dyeing process provide the correct ionic strength (as electrolyte), pH (acids and bases), and moderation of the dye-fiber-water interactions (often surfactants). While it may not be obvious, similar interactions take place in textile printing with dyes, albeit in the confines of a small volume of print paste, which will also include substances (“thickeners”) to control its rheology and balance the ready transfer of paste onto the fabric with the need to maintain a sharp printed “mark.” These various dyeing and printing adjuncts, however, do not fall under the definition of mordants. They remain within the exhausted dyebath or print paste at the end of the process and are lost from the colored fiber in any subsequent rinsing.

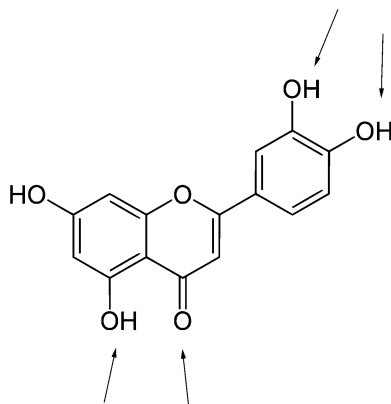
Natural Fibers

As mentioned earlier, the use of mordants was and is largely confined to the use of natural fibers, which fall into two distinct groups: protein (wool and silk) and cellulosic (chiefly cotton and linen). Mordants behave somewhat differently for each of these two groups of fibers. Protein fibers comprise many different amino acids with a range of functional groups that provide binding opportunities for colorants of all kinds [6]. Natural dye users overwhelmingly prefer to dye wool for this reason. Mordants are/were nevertheless widely used to provide better fastness and to broaden the range of colors obtainable from a limited number of natural dyes. Cellulose lacks this inherent affinity for natural colorants, and with the notable exceptions of indigo (which does have substantivity, albeit low), turmeric, archil, saffron, and annatto (which have fastness limitations), natural dyes do not readily dye cellulose. Mordants provide the main means by which cellulose may be satisfactorily dyed with natural dyes.

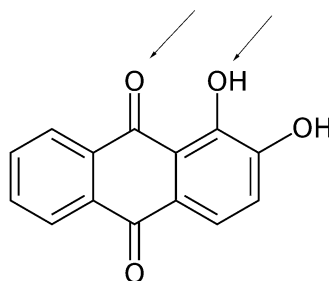
Natural Dyes

Many natural organic substances are colored, and a few of those colors can be employed as dyes. Chemically, many are flavonoid derivatives (Fig. 1), with anthraquinones (Fig. 2), carotenoids, and indigoids also providing useful dyes [7, 8]. Most provide yellow-orange-brown colors; bright reds are less common, and blues are rare. Fastness varies, especially to light, and is influenced by the choice of mordant and efficiency of fiber-mordant-dye binding.

Over time, expert European dyers learned how to apply these dyes with the appropriate mordants to give bright and fast colors, versus those that were less satisfactory. Thus, Venetian dyers were of either greater or lesser arts. French dyers were “au grand teint” and “au petit teint” [9]. In each case, the former were the more expert. Global exploration would eventually provide better natural dyes. Woad was the original blue in Europe,



Coloration, Mordants, Fig. 1 Luteolin, the main coloring matter of weld (“dyer’s rocket”). Potential binding sites for mordants are arrowed



Coloration, Mordants, Fig. 2 Alizarin, the main coloring matter of madder. Potential binding sites for mordants are arrowed

replaced with indigo (the same essential colorant) when that became available. Reds and purples were from madder and kermes, the latter replaced when cochineal from Central America was discovered. Yellows were from weld or Persian berries, later from quercitron found in North America. Dye woods (sappan, Brazil, cutch, fustic, logwood) were also important. Other than woad or indigo, these colorants were used in conjunction with a mordant. The inventive chemistry that allowed these natural dyes to give good results when they were all that was available, today, allows craft dyers and researchers to reexamine a wide range of plants as sources for natural dyes in the supposition that these dyes are somehow more “sustainable.”

Mordants for Cotton and Cellulosic Fibers

Since few natural dyes display substantivity toward cellulosic fibers, the use of mordants is essential to achieving satisfactory dyeings and prints with those dyes. Two sources [10, 11] give a valuable survey of the use of mordants on cellulosic substrates when such mordants were being widely used. The most effective mordants were metallic salts that could be applied as a soluble salt, rendered less soluble, and subsequently undergo complex formation with the dye. For dyeing a solid color, the fabric would be padded with this metallic salt. For printed designs, the metallic salt would be incorporated into a print paste. Such a metal salt, if it remained soluble, would redissolve in the subsequent dyebath: the dye-mordant complex formed in the bath would tend to deposit on the textile surface and later be prone to rub off. Thus, some means of reducing the solubility of the metal ion and keeping it in place on the substrate as the dye-mordant reaction took place in the dyebath were required, especially in a printed substrate where reserve of the unprinted (white) areas was essential [1].

Insolubilization could be accomplished in a number of ways. Using aluminum as an example:

1. A soluble aluminum salt such as aluminum sulfate or aluminum potassium sulfate (potassium alum) could be neutralized with carbonate or bicarbonate to produce a “basic aluminum sulfate” or “basic alum,” respectively. Padding with this, followed by treatment with ammonia, would precipitate aluminum hydroxide on the fiber. Alternatively, the mordant metal could be precipitated as the corresponding phosphate, arsenate, or silicate [10].
2. Co-application of a complex anionic substance would provide insoluble complex salts. Tannins from a range of sources were ubiquitous in this regard, from galls, myrbolam, cutch, or (later) purified tannic acid, resulting in a deposit of aluminum tannate [10]. The effectiveness of “mordants” based on modified vegetable oils (usually castor or olive oil) was

probably due to a similar mode of action: Turkey red dyeing is discussed below. (Note: cotton will absorb tannic acid from solution, and cotton so treated does dye more readily, so tannic acid does have a moderate effect as a mordant by itself.)

3. Salts such as aluminum acetate could be “aged” in moist conditions under which acetic acid would be lost, generating an insoluble metal hydroxide. A “dunging” (originally with cow dung, later with synthetic alternatives such as sodium phosphate or arsenate) process would both complete the fixation of the mordant and remove unfixed mordant. This process was especially important in printing [11].

Given the above example, it is not surprising that nineteenth-century references to mordants usually include a similar application of a metal salt that would subsequently be reacted with a second metal salt to generate a colored inorganic compound. For example, a “lead mordant” (lead acetate) applied by padding and drying would be passed through a bath of sodium or potassium dichromate to produce insoluble yellow lead chromate [12]. This precipitation reaction is sufficiently rapid that the insolubilization of the original “mordant” was not required. This use of the word mordant is not current.

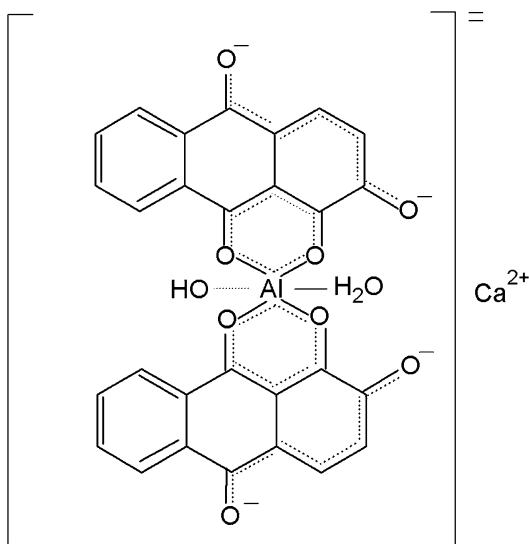
Excluding, therefore, the use of the word mordant as a component of a precipitated mineral colorant, aluminum and iron were used most extensively as mordants for natural dyes on cotton in both dyeing and printing. Stannic salts were occasionally used. Both iron and aluminum could be produced as the acetate. Iron could be directly dissolved in pyroligneous (acetic) acid, while a reaction between aluminum or ferrous sulfate and lead acetate or calcium acetate would generate a lead or calcium sulfate precipitate and leave aluminum or ferrous acetate in the supernatant liquid. These acetate solutions were referred to as purple liquor (for iron) and red liquor (for aluminum), based on the colors they would produce in a madder containing dyebath [13]. Tin salts were common adjuncts to aluminum salts, where they prevented dulling by the presence of small amounts of iron salts. Chromium salts were

less easy to precipitate in the fiber and were not extensively used, other than to provide a black color with logwood in which processed dye and mordant were often applied together (compare the dyeing of wool by the metachrome method discussed below). Texts available during that time also include a range of other metal salts among the discussions of mordants for cotton but indicate that their function is often that of an oxidizing (Cu, Cr) or reducing (Sn) agent.

The immersion of a mordanted fabric in a dyebath would allow the formation of a dye-metal complex. (This is discussed in greater detail under mordant dyeing of wool, below.) In a fabric print, the unmordanted dye in the dyebath would inevitably stain the unmordanted portions of the fabric, and the fastness of the dye-mordant complex in the colored areas is indicated by its ability to survive the subsequent “clearing” of dye from white parts of the printed fabric by extensive washing processes.

As outlined above, the process is reasonably straightforward, although requiring care and attention to achieve successfully. Some idea of the added complexities and obfuscation generated by ignorance can be gleaned from any reading of the history of Turkey red dyeings. Modern science has identified the final product on the fabric as a 2:1 alizarin/aluminum complex anion with an associated calcium cation [14] (Fig. 3). But for years, the process was shrouded in mystery, and much money was paid for supposed recipes for this color [13]. As originally described, the process involved many steps carried out over a period of weeks or months. The deposit of fatty acids (from olive or castor oil) was a part of the process, leading to references to “oil mordants.” Advances during the nineteenth century such as the introduction of synthetic alizarin, “Turkey red oil,” and chemical bleaching methods allowed for the dyeing to be carried out in days instead of months [15]. The structure of Turkey red oil and its role as part of the mordanting process have recently been (re)examined [24].

Given the difficulty in dyeing cotton and other cellulosic fibers with natural dyes, it is not surprising that as scientific knowledge increased in the early nineteenth century, many attempts were



Coloration, Mordants, Fig. 3 “Turkey red” (alizarin-aluminum complex). Compare Fig. 2, alizarin.

made to modify those fibers to improve their dyeability. Dyers of the nineteenth century attempted such modifications to cotton. The substances have been described as “mordants” [11], but if these modifications are carried out in bulk, separately from the dyeing process, they stretch the definition somewhat. The modifications were presumably prompted by the greater substantivity of dyes for protein fibers and mostly consisted of depositing protein material onto cotton. Albumen from eggs or blood, gelatin, and casein and lactarine from milk were used in these attempts. The efforts were classified as “animalizing” cotton. Such modifications continue to this day: cotton can be pretreated to have cationic groups and readily takes up anionic dyes [16]. Cotton so treated is not referred to as being mordanted, however.

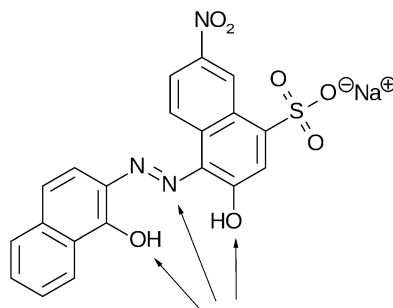
A new category of mordants for cotton was required when the first synthetic dyes were introduced following Perkin’s synthesis of mauveine in 1856. For almost 30 years, many of these new dyes comprised what are now classified as acid and basic dyes, which have no substantivity toward cotton. Dyers were obviously eager to make use of the bright colors they offered, and the precipitation of tannic acid with salts of antimony (in the form of antimony potassium tartrate,

or “tartar emetic”) on the fiber provided an anionic substrate with which cationic basic dyes would interact sufficiently well to provide substantivity and fastness [15].

A subplot in the saga of cotton and mordants occurs in efforts in the twentieth century to improve the fastness of the direct dyes. These efforts comprised aftertreatments to a dyed material. Mordants typically are applied before the application of dye; these aftertreatments are not usually regarded as mordants, but the interactions have factors in common with those that dyes and mordants undergo. For example, anionic direct dyes for cotton could be aftertreated with cationic compounds to form a salt complex with reduced solubility and better fastness to washing. Direct dyes with functional groups capable of chelating metal ions might be aftertreated with copper or chromium salts: a dulling of shade would be compensated by improved fastness to light and (less so) to washing. This technique is technically indistinguishable from the application of mordant dyes to wool by afterchroming but is not regarded as an example of the use of mordants [3].

The same (natural) dye will provide distinctly different colors depending on which metallic salt it interacts with. The functional groups on the polypeptide chains that form the keratin of which wool is composed include groups, notably carboxylic acid, with which metal ions can form bonds, holding them in place during subsequent interaction with the dye. Thus, the insolubilization required for cotton mordants is not needed. For those reasons, the number of metal salts useful as mordants is greater for wool than for cotton and includes tin, copper, and chromium. Wool additionally has sulfur-containing amino acids that will reduce (or maintain in a reduced form) polyvalent metals: thus, chromium is applied as dichromate (Cr^{6+}) and interacts with dye as Cr^{3+} . Tin is applied and interacts as Sn^{2+} . Dye recipes often include mild reducing agents (tartrate, oxalate) to assist this.

Natural dye structures (Figs. 1 and 2) have quinonoid, hydroxyl, and carboxy groups capable of chelating metal ions. Synthetic mordant dyes have the same groups plus nitrogen-containing (amino and azo) groups (Fig. 4). Dyes are usually



Coloration, Mordants, Fig. 4 CI Mordant Black 11 (14645). Sites for coordination with chrome mordant arrowed

bi- or tridentate ligands, and the metals they interact with usually have coordination numbers of four or six.

Mordant Dyeing of Wool

As discussed earlier, the use of a mordant for applying natural dyes to wool has less to do with achieving substantivity than it has to providing better fastness than the unmordanted dye and to extending the range of colors available. Mordants for wool were (and are) overwhelmingly metallic compounds: no parallel with the animalizing reactions on cotton discussed above was developed. The non-mordant cationization of cotton is perhaps echoed in comparatively recent research to render wool dyeable with disperse dyes [17].

The chelation reactions on the fiber between chromium and synthetic mordant dyes have been the most widely studied, but the principles involved are the same as those between other metals and natural dyes. Dye makers simplified the mordant dyeing process by carrying out the dye-metal binding during dye manufacture and selling “premetallized” dyes (also known as metal-complex acid dyes), and the chemistry of this complexation is more easily studied in the absence of the fiber [2, 18, 19]. While these are overwhelming complexes involving chromium and cobalt, iron and aluminum have been examined as less polluting alternatives, and copper-complex direct and reactive dyes are used on cotton. Dye-metal (1:1) complexes were

introduced in the early twentieth century, with 2:1 dye-metal complexes introduced in mid-century. Chromium has a coordination number of six: the dye molecule in 1:1 chromium complexes provides three ligand interactions (binding sites), and the remaining three ligand sites are occupied typically by water. It has been suggested that these sites can interact with groups on the wool fiber, although evidence is lacking. If they do so, however, this would represent one of the few instances where the common metal model of a mordant as a bridge linking dye and fiber is relevant. The metal atom in a 2:1 complex has no such spare binding capacity and achieves fastness via large molecular size and low solubility. In theory, either 1:1 or 2:1 complexes might be formed by a mordant dyeing process, but evidence suggests that the 2:1 complexes predominate. Whether produced on the fiber or as a metal-complex dye, the resulting complex is more stable to light and has limited solubility and thus good fastness to wet treatments. The traditional method of mordant dyeing of (natural) dyes on wool involved pretreating the wool with a solution of the metal salt and subsequently introducing the mordanted wool into a dyebath. This sequence is still used for dyeing natural dyes on wool today. When synthetic (mordant) dyes were developed, their varied chemistry could generate the range of shades required, and the need for multiple metal mordants declined, with chromium (as dichromate) providing the best fastness, and this became the main choice for mordant dyeing using the “chrome mordant method” [20, 21]. Interest in making the processes more efficient took advantage of the independent substantivity of both dye and mordant. They could be applied together (the “metachrome method”) in a manner reminiscent of the chromium-logwood black dyeing of cotton. While this is a one-bath process and seems an obviously more efficient means of application, it is limited to dyes that exhaust well at neutral pH, are not precipitated by dichromate, and do not reduce Cr^{6+} to Cr^{3+} [20]. Eventually, the process in which the dye was applied first and subsequently treated in a bath of dichromate ions became most widely used as the “afterchrome” or “topchrome” method. As discussed above,

this is technically indistinguishable from the after-treatment of a direct dyeing on cotton with a copper (Cu^{2+}) salt.

Synthetic Fibers

Given the similar dyeability of nylon to wool and the search for fastness in nylon dyeings, it is not surprising that chrome mordant dyes have been applied to nylon. The lack of reducing groups in nylon means that a reducing agent is required in the process to convert Cr^{6+} to Cr^{3+} [22]. Chromium-complexed dyes are rarely applied to nylon today. In an interesting twist, the tannic acid/antimony mordant for basic dyes on cotton has been applied as an aftertreatment (“back-tanning”) to improve the fastness of acid dyes on nylon, although no interaction with the dye is suggested.

Acrylic fibers, introduced in the 1950s, were initially difficult to dye in dark shades with acceptable fastness. Before the development of modified basic dyes, an early solution was the so-called cuprous ion method in which the fiber was treated with a solution of Cu^+ ions, and the fibers were then dyeable to dark shades with acid and direct dyes. A number of variations were published, including the application of dye and copper ions simultaneously and the use of reducing agents, but the process was superseded before becoming a standard dyeing method. This is clearly a mordant dyeing technique, although none of the descriptions of the time refer to it as such [23].

Conclusion

Mordants have a long history in enhancing the coloration of textile fibers. They are essential in the dyeing of cellulosic fibers with natural and basic dyes and improve the fastness and color range obtainable on wool with natural dyes. With the advent of synthetic dyes, their use became restricted to the chrome dyeing of wool, but they have found little value in any dyeing of synthetic fibers. They are rarely used in large-

scale commercial dyeing today but are widely used among home and craft dyers for whom the use of natural dyes remains popular.

While the details may sometimes be hard to unravel from historical or unscientific descriptions/interpretations, the majority of mordants are metal ions (most notably of Cr, Al, Fe, Sn, Cu) that react with dyes to form coordination complexes of low solubility and better fastness. Other useful mordants are those that interact ionically with dyes, again forming a lower solubility complex.

Cross-References

- [Coloration, Fastness](#)
- [Coloration, Textile](#)

References

1. O'Neill, C.: *A Dictionary of Dyeing and Calico Printing*. Baird, Philadelphia (1869)
2. Shore, J.: *Colorants and Auxiliaries*, 2nd edn. Society of Dyers and Colourists, Bradford (2002)
3. Society of Dyers and Colourists, American Association of Textile Chemists and Colorists: *Colour Index*. Society of Dyers and Colourists, Bradford. American Association of Textile Chemists and Colorists, Lowell (1956)
4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Bancroft, E.: *Experimental Researches Concerning the Philosophy of Permanent Colours and the Best Means of Producing them*. Cadell and Davies, London (1813)
6. Pailthorpe, M.: The theoretical basis for wool dyeing. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 52–87. Society of Dyers and Colourists, Bradford (1992)
7. Hofenk-de Graaff, J.H.: *The Colorful Past. Origins, Chemistry and Identification of Natural Dyestuffs*. Archetype, London (2004)
8. Lee, D.: *Nature's Palette: The Science of Plant Color*. University of Chicago, Chicago (2007)
9. Brunello, F.: *The Art of Dyeing in the History of Mankind*. NeriPozza, Venice (1973)
10. Knecht, E., Fothergill, J.B.: Mordants, etc. In: Knecht, E., Fothergill, J.B. (eds.) *The Principles and Practice of Textile Printing*, 4th edn, pp. 203–269. Griffin, London (1952)
11. Hummel, J.J.: Mordants. In: Hummel, J.J. (ed.) *The Dyeing of Textile Fabrics*, pp. 156–247. Cassel, London (1885)
12. Sansone, A.: The Printing of Cotton Fabrics Comprising Calico Bleaching, pp. 129–130. *Printing and Dyeing*, Heywood (1887)
13. Bide, M.J.: Secrets of the printer's palette. In: Welters, L.M., Ordonez, M.T. (eds.) *Down by the Old Mill Stream: Quilts in Rhode Island*, pp. 83–121. Kent State University, Kent (2000)
14. Gordon, P.F., Gregory, P.: *Organic Chemistry in Colour*. Springer-Verlag, Berlin (1983)
15. Matthews, J.M.: *Application of Dyestuffs*. Wiley, New York (1920)
16. Lewis, D.M., McIlroy, K.A.: The chemical modification of cellulosic fibers to enhance dye-ability. *Rev. Prog. Color.* **27**, 5–17 (1997)
17. Lewis, D.M., Pailthorpe, M.T.: The modification of wool with reactive hydrophobes. *J. Soc. Dye. Colour.* **100**, 56–63 (1984)
18. Burkinshaw, S.M.: Dyeing wool with metal-complex dyes. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 196–221. Society of Dyers and Colourists, Bradford (1992)
19. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*. VCH, New York (1987)
20. Bird, C.L.: *The Theory and Practice of Wool Dyeing*, 3rd edn. Society of Dyers and Colourists, Bradford (1963)
21. Duffield, P.: Dyeing wool with acid and chrome dyes. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 176–195. Society of Dyers and Colourists, Bradford (1992)
22. Burkinshaw, S.M.: *Chemical Principles of Synthetic Fibre Dyeing*. Blackie, London (1995)
23. Wilcock, C.C., Ashworth, J.L.: *Whittaker's Dyeing with Coal Tar Dyestuffs*, 6th edn. Balliere, Tindall and Cox, London (1964)
24. Wertz, J.L., France, D.J., Quye, A.: Spectroscopic analysis of Turkey red oil samples as a basis for understanding historical dyed textiles. *Color. Technol.* **134**, 319–326 (2018)

Coloration, Textile

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

Synonyms

[Colouring](#); [Coloring](#); [Dyeing](#); [Mass pigmenting](#); [Printing](#)

Definition

Dyeing can be described as the uniform application of colorant(s) to a coloring medium. The coloring of textiles may involve mass pigmenting (involving compounding), dyeing, and printing processes. The coloring medium in textile dyeing may take different physical forms (such as loose fiber, yarn, tow, top, woven, nonwoven and knitted substrates in open width or rope form), whereas in printing colorants are added to selected regions of the medium which is usually in a fabric form. Dyeing of homogenous fibers should result in a uniform solid color. Multicolored effects may be obtained by dyeing a blend of different fibers, bi- or multicomponent synthetic fibers, or via multiple colorant or illuminated discharge printing. In a technique known as space dyeing, colorant(s) can be applied from nozzles that inject different dyes to yarns and force steam in the vessel, thus generating a multicolored pattern.

There are difficulties encountered in controlling the physicochemical changes that occur during dyeing when attempting to maximize color yield, levelness of dyeing, color fastness, etc. Therefore, a vast sub-technology of specialty chemical auxiliaries is used in preparation for dyeing and in the dyeing process itself, such as levelling agents, dispersing agents, antifoams, etc., as well as auxiliaries specifically designed for aftertreatments. Moreover, from the practical standpoint, methods of applying colorants to the coloring medium can be broadly divided into batch, semicontinuous, and continuous processes. These are discussed in more detail in the following sections.

Dyeing Process

The science of coloration entails several domains including chemistry, physics, physical chemistry, mechanics, fluid mechanics, thermodynamics, and others. Devising the most efficient dyeing process typically involves several concerns including machinery design, preselection of dyes of compatible properties, use of pH versus time profiles, selection of liquor ratio, flow rate and

flow direction reversal times, and design of temperature versus time profiles. It is thus clear that dyeing is a complicated process with many different phenomena occurring simultaneously that require a suitable level of control.

When a textile fiber comes in contact with a medium containing a suitable dye under suitable conditions, the fiber becomes colored, the color of the medium decreases and that of the fiber increases, thus resulting in the dyeing of the substrate. True dyeing occurs when the dye is absorbed with a decrease in the concentration of dye in the dyebath and when the resulting dyed material possesses some resistance to the removal of dye by washing [1], rubbing, light, and other agencies.

Coloration of textiles is not limited to the simple impregnation of the textile fiber with the dye that occurs during the initial phase of the dyeing process. A dye is taken up by a fiber as a result of the chemical or physical interactions between the colorant and the substrate. Many dyes for textiles are water soluble, and their molecules are dissociated into positively and negatively charged ions in aqueous environments. The uptake of the dye by the fiber will depend not only on the nature of the dye and its chemical constitution but also on the structure and morphology of the fiber.

Textile Fibers

A fiber is characterized by its high ratio of length to thickness and by its strength and flexibility. Fibers may be of natural origin or formed from natural or synthetic polymers. They are available in a variety of forms. Staple fibers are short, with length-to-thickness ratios around 10^3 – 10^4 , whereas this ratio for continuous filaments is at least several millions [2]. The form and properties of a natural fiber such as cotton are fixed, but for man-made fibers, a wide choice of properties is available by design. The many variations include staple fibers of any length, single continuous filaments (monofilaments), or yarns comprised of many filaments (multifilaments). The fibers or filaments may be lustrous, dull or semi-dull, coarse, fine or ultra-fine, circular or of any other

cross section, straight or crimped, regular or chemically modified, or solid or hollow.

From a processing standpoint, natural fibers have a number of inherent disadvantages. They exhibit large variations in staple length, fineness, shape, crimp, and other physical properties, depending upon the location and conditions of growth. Animal and vegetable fibers also contain considerable and variable amounts of impurities which must be removed prior to commencing dyeing. Man-made fibers are much more uniform in their physical characteristics. Their only contaminants are small amounts of slightly soluble low M_w polymer and some surface lubricants and other chemicals added to facilitate processing. These are relatively easy to remove compared to the difficulty of purifying natural fibers.

Water absorption is one of the key properties of a textile fiber that influences their coloration. Protein or cellulosic fibers are hydrophilic and absorb large amounts of water, which causes radial swelling. Hydrophobic synthetic fibers, such as polyester, however, absorb almost no water and do not swell. The hydrophilic or hydrophobic character of a fiber influences the types of dyes that it will absorb. The ability to be dyed to a wide range of hues and depths is a key requirement for almost all textile materials.

Another important property of a textile fiber is its moisture regain, which is the mass of water absorbed per unit mass of completely dry fiber, when it is in equilibrium with the surrounding air, at a given temperature and relative humidity. The regain increases with increase in the relative humidity but diminishes with increase in the air temperature. Water absorption by a fiber liberates heat (exothermic) and will therefore be less favorable at higher temperatures. The heat released is often a consequence of the formation of hydrogen bonds between water molecules and appropriate groups in the fiber. When the final regain is approached by drying wet swollen fibers, rather than by water absorption by dry fibers, the regain is higher. For hydrophilic fibers such as wool, cotton, and viscose, the relatively high regain values significantly influence the gross mass of a given amount of fiber. This is significant in dyeing. Amounts of dyes used are usually expressed as a

percentage of the mass of material to be colored. Thus, a 1.0% dyeing corresponds to 1.0 g of dye for every 100 g of fiber, usually weighed under ambient conditions. For hydrophilic fibers, the variation of fiber mass with varying atmospheric conditions is therefore an important factor influencing color reproducibility in repeat dyeings.

Stages of Coloration Process

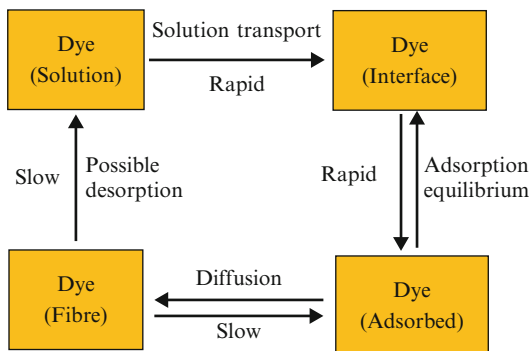
Most textile dyeing processes initially involve transfer of the colored compound, or its precursor, from the aqueous solution onto the fiber surface, a process called *adsorption*. From there, the dye may slowly *diffuse* into the fiber. This occurs through pores or between fiber's polymeric chains, depending on the internal structure of the fiber. The overall process of adsorption and penetration of the dye into the fiber is called *absorption*. Absorption is a reversible process. The dye can therefore return to the aqueous medium from the dyed material during washing, a process called *desorption*. Besides direct absorption, coloration of a fiber may also involve precipitation of a dye inside the fiber, or its chemical reaction with the fiber. These two types of processes result in better fastness to washing, because they are essentially irreversible.

Dye Transport from the Bulk Solution to the Fiber Surface

Dyeing is a process which takes time. The transfer of a dye molecule from the dye solution into a fiber is usually considered to involve the initial mass transfer from the bulk dye solution to the fiber surface, adsorption of the dye on the fiber surface, followed by diffusion of the dye into the fiber as depicted in Fig. 1.

When a fibrous assembly is immersed in a dye solution, the rate at which the dye is taken up is generally dependent upon the extent to which the liquor is agitated and tends to approach a maximum value when the stirring is vigorous.

The transfer of dye from the bulk solution to fiber surface is fast, and the rate generally increases



Coloration, Textile, Fig. 1 Dye transfer from bulk solution into a fiber [3]

with increasing the flow rate. The adsorption equilibrium is also rapid, so it is usually assumed that the overall rate of dyeing depends on the rate of diffusion of the dye into the fiber. Inadequate control of the rate of dye adsorption will result in unlevel dyeings unless the dye can subsequently migrate from deeply dyed to lightly dyed regions of the substrate [1]. Therefore, the control of the first two stages of the process, namely, the initial mass transfer from the bulk solution to the fiber surface and adsorption of the dye on the surface, is important for a level dye distribution throughout the substrate to be achieved.

A fundamental issue in dyeing is to ensure that the dye liquor penetrates all parts of every fiber and is distributed within the substrate as evenly as possible. In practice, however, there may be differences in fibers, to a greater or lesser extent, because of natural or process-related parameters. These variations are of major importance, since, in wet processes, changes in regional density of the substrate result in variations in the degree of dye penetration and differences in flow rate which can lead to shade differences in dyeing.

The influence of agitation in increasing the rate of dye uptake is dependent in large measure on the hydrodynamic complexity of the system. Unfortunately, flow through the substrate cannot be described in any simple fashion with fibrous assemblies, due to the extreme complexity of defining the flow of liquor through a mat of fibers or through yarns or cloth.

In spite of the complications of real systems, some of the principles governing the process may be elucidated by consideration of simple ones, e.g., a plane sheet or film of material immersed in dye liquor whose direction of flow is parallel to the sheet. On making contact with the dye liquor, dye is adsorbed by the film so that the neighboring liquor becomes deficient in dye; dye is transported to the surface by dispersion from the bulk, but the quantity transferred is modified by the speed at which the liquor passes through the film.

Calculations of the rate at which the fibers can take up dye require knowledge of the flow pattern of the liquor before analysis of diffusion may be attempted. Experimental investigations of flow of dye liquor through masses of fiber lead to the conclusion that, at the common rates used in dyeing, the flow is streamline rather than turbulent, so attention may be confined to streamline conditions.

Intermolecular Forces Operating Between Colorants and Textile Material

The strongest dye-fiber attachment is that of a covalent bond. Another important interaction between dyes and fibers includes electrostatic attraction, which occurs when the dye ion and the fiber have opposite charges. Hydrogen bonds may also be formed between a specific range of colorants and textile fibers. In addition, in nearly all dyeing processes, van der Waals forces and hydrophobic interactions are involved. The combined strength of the molecular interactions is referred to as the *affinity* of the dye for the substrate. The substantivity of the dye is a less specific term and is often used to indicate the level of exhaustion (which is described in the following sections) [4]. Thus, substantivity is the attraction between the substrate and a dye under precise conditions where the dye is selectively extracted from the medium by the substrate. Different types of textile fibers require different kinds of dyes, and in general, dyes which are suitable for one type of fiber may not dye other types effectively.

Levelness

Levelness is the uniformity of dye distribution (and hence color) on textiles. Two fundamental mechanisms contribute to a level dyeing. One is the initial sorption of dye during the dyeing; the other is the migration of dye after initial sorption on the fiber. An initial level sorption will lead to a level dyeing. An unlevel sorption may be corrected if sufficient migration takes place. These mechanisms are affected by dyes and chemicals, by textile substrate, and by controlling the parameters of the dyeing process such as dyebath pH, liquor ratio, flow rate, and temperature. Some dyes are more likely to level out, especially if they are small and do not have a high degree of affinity towards the substrate; other dyes on the other hand tend to be much less likely to migrate. These are often dyes of large size or with strong affinity towards the substrate.

Dyeing of Various Textile Materials

Textile fibers can be dyed at various stages of production such as loose fiber, top, tow, yarn, fabric, or garment. The levelness requirements for dyeing loose fibers are less strict than for dyeing at the yarn stage or at later stages, since further processing of the loose fiber results in some mixing of the dyed fibers, e.g., during carding or gilling, which improves the levelness of the resultant color. For yarn dyeing levelness is more critical because while the yarn will be made into fabric, either by knitting or weaving, or used in the construction of carpets, any unevenness in the dyeing of the yarn will show in the finished goods. In the case of fabric and especially garment dyeing, the process will be less forgiving and variations will have to be remedied in different ways, either by over dyeing the substrate to a darker shade or by stripping the color and reapplying the colorant. Dyeing textiles at earlier stages of production such as loose fiber also reduces the degree of flexibility in response to market needs. Garment-dyed material can reach the market quickly and meet the rapidly changing

demands of the market, whereas in the case of fibers or yarns, typically several additional processes have to be completed before the dyed material reaches the consumer. In the case of substrates with striped patterns yarn dyeing is quite common. When blending dyed yarn with undyed yarn to produce a woven pattern, it must be considered that subsequent bleaching and scouring may be needed. Thus, dyes of suitable fastness properties, to bleach and various agencies, have to be employed to ensure no staining of adjacent white regions occurs. Dyeing of fabrics and garments with variations in fabric density or stress can also result in systematic unlevelness. A common type of unlevelness in dyeing of nylon fabrics is known as Barré which appears as stripes across the substrate. A specific visual appearance test method to rate the degree of Barriness of dyed substrate has been introduced over the years.

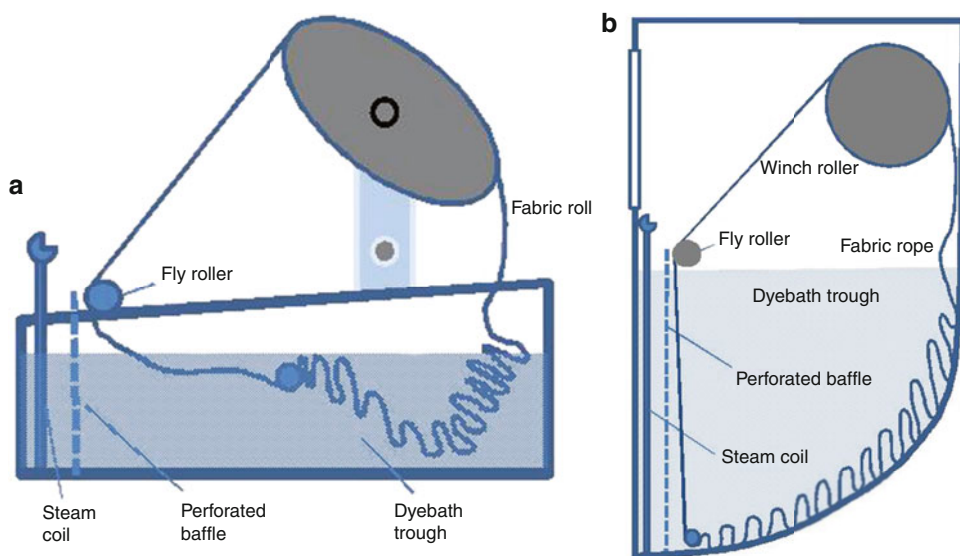
Batch, Semicontinuous, and Continuous Coloring Processes

There are three main types of processes for the dyeing of textile materials: batch, continuous, and semicontinuous. Batch processes are the most common method used to dye textile materials and often depend on the type of equipment available and the weights or lengths of the material to be dyed. Batch dyeing is often called exhaust dyeing because the dye is gradually transferred from a relatively large volume dyebath to the material being dyed over a long period of time. Batch dyeing, therefore, involves applying a dye from a solution or a suspension at a specific ratio of liquor to textile substrates where the depth of the color obtained is mainly determined by the amount of colorant present in relation to the quantity of fiber (known as L:R or liquor to goods ratio), although other factors can also influence the overall dye uptake. Batch processes are thus designed for specific quantities of substrate from few grams to several hundred kilograms. Batch dyeing of most natural fibers is carried out under atmospheric pressure conditions. Operations may also be carried out under elevated pressures. For

instance, in the batch dyeing of polyester substrates, high temperatures (HT), around 125–130 °C, and correspondingly elevated pressures are required if the use of carriers is to be avoided. High-pressure beams, jets, and jigs can be used. Although such equipments tend to be more expensive than conventional atmospheric pressure equipments, their cost is more than offset by the omission of carriers. At high temperatures the diffusion rate of the dyes is high enough to produce satisfactory shades in a dyeing time of about one hour. Batch dyeing of polyester, however, can give rise to some bruises and pillings due to hydraulic and mechanical impacts during high temperature dyeing process. Generally, flexibility in color selection in batch dyeing is high, but the cost of dyeing is lower the closer the dye application is to the end of the manufacturing process for a textile product. Two examples of a common type of batch dyeing machine known as beck or winch (shallow-draft and deep-draft winches) are given in Fig. 2.

Semicontinuous dyeing is carried out in a continuous range where the substrate is fed into the dyeing range from one end and collected at the other. In semicontinuous processes typically fixation and washing steps are carried out

discontinuously. Pad rolls transfer the colorant that is picked up from a trough into the substrate, and the process is based on impregnation of the substrate followed by fixation of the colorant. The pressure applied by rolls to impregnate the substrate can be adjusted to squeeze the excess liquor out of the substrate and obtain the required wet pickup percentage. Since the initial uniformity of dye deposition on the substrate is critical for a level dyeing, padding rolls responsible for the transfer of the colorant onto the substrate must have a uniform surface with no indentations. The pressure across the rolls should also be adjusted and controlled regularly to ensure the uniformity of color transfer onto the substrate. The dyeing range may include cans or alternative forms of dryers to prevent migration of dye across the substrate prior to fixation. At the end of the range, the substrate may be rolled onto a beam, covered with plastic sheets, and kept overnight, to enable fixation of the dye, or transported to other sections for subsequent fixation, washing, and aftertreatment operations. Pad-batch dyeing is a specific type of semicontinuous coloration process which is common in the application of reactive dyes to cellulosic substrates. In this process the batch is left overnight to enable the colorant to



Coloration, Textile, Fig. 2 (a) Shallow-draft winch dyeing machine. (b) Deep-draft winch dyeing machine

react with the substrate. Other forms of pad-fix processes may include pad-bake and pad-steam.

Continuous dyeing refers to operations at constant composition involving several application and wash boxes (troughs), where a long length of textile fabric is pulled through each stage of the dyeing process including fixation and aftertreatment. Continuous dyeing operations are common when dyeing large quantities of substrate. This is often carried out on cotton and its blends with synthetic fibers such as polyester. In continuous dyeing processes, fixation and subsequent wash and rinse operations are combined with the coloration process to enable rapid throughputs. Thermofixation (commercially known as Thermosol) which is commonly carried out on polyester and polyester blends is particularly suited to continuous dyeing processes, since it involves padding the dyestuff from dispersion onto the fabric, drying, and then heating the padded substrate to a temperature of 180–220 °C. A specific category of disperse dyes capable of sublimation is employed for this purpose. At such temperatures, diffusion rates of sublimated dyes are so high that a few seconds suffice for adequate penetration of dye molecules into the substrate. Several variables, however, affect the final shade of the dyed fabric during the thermofixation process. Some of these variables are Thermosol period, temperature, type of disperse dyestuff, and pad bath auxiliaries.

In continuous printing processes, colorants are applied to specific sections of the cloth using a number of techniques that may include roller, flatbed, and rotary screen printing systems to obtain a preset design. Dye fixation is carried out by steaming or baking the printed material followed by washing to remove surplus dye, thickeners and any other auxiliaries.

The details of the dyeing process can vary considerably between different types of textile materials employing different types of dyeing equipment. For example, the maximum permissible rate at which the temperature of the bath can be raised may be determined by the relationship between the rate of circulation of the dye liquor and the rate of transfer of the dye from bath to fiber such as in the dyeing of yarn in hank dyeing machines.

The criteria for choosing a dyeing process vary and may include the following:

- Shade range
- Fastness requirements
- Quality requirements and control
- Cost
- Equipment availability
- Dye selection

The aim of a successful dyeing process is to achieve the desired shade, at the right price, with sufficient levelness, whether dyeing loose fiber, yarn, or piece goods, with sufficient color fastness to withstand both processing and consumer demands, but without adversely affecting the fiber quality. Of these, an acceptable level of uniform dye uptake at all parts of the substrate may be the most important criterion.

A typical dyeing process may be divided into several steps as follows:

- Establishment of equilibrium between associated molecular dye and single molecules of dye in solution
- Diffusion of monomolecular dye to the diffusional boundary layer at the fiber surface
- Diffusion of dye through the boundary layer at the fiber surface
- Adsorption of dye at the fiber surface
- Diffusion of the dye into the fiber interior
- Desorption and readsorption of dyes (migration)

These steps form a reversible equilibrium system. Each of the six steps can influence the levelness of dyeings. A complete quantitative analysis of the effects of many factors which influence the levelness of dyeing would require the development of a mathematical model involving a significant number of parameters. This is, however, a very difficult task. For practical applications, one may initially try and identify a few variables, which are thought to have a larger impact on levelness and restrict the development of the model to the effects of these few most important variables.

The Donnan equation involves nine factors: the concentration of dye in fiber, the concentration

of dye applied, the liquor ratio, the distribution of ions between solution and fiber, the ionic charge on the dyestuff molecule, the internal volume of the fiber, the affinity of the dye, the gas constant, and the dyeing temperature. All together, these terms describe the dyebath conditions.

To obtain reproducible dyeings, whether this is on a laboratory, pilot plant, or bulk scale, the following factors in the dyeing process must be controlled or measured:

- Quality of water supply
- Preparation of substrate
- Dyeability of substrate
- Weight of substrate
- Moisture content of substrate at weighing
- Selection of dyes
- Standardization of dyes
- Weighing of dyes and chemicals
- Dispensing method for dyes and chemicals
- Moisture content of dyes
- Liquor to goods ratio
- Dyebath additives
- pH of dyebath
- Machine flow and reversal sequence
- Time/temperature profile

Various workers have placed different degrees of emphasis on each of the above factors and also on the factors which are not listed above. There may also be disagreement on the mechanism by which any one of these factors operates. Furthermore, it should be noted that the parameters in this list are not quite independent, and in some cases the effect of two of these factors may be considered as one effect.

Exhaustion

Dyeing is carried out either as a batch exhaustion process or as a continuous impregnation and fixation process. In exhaust dyeing, all the material is in contact with all the dye liquor from where the fibers absorb the dyes. The dye concentration in the bath therefore gradually decreases. The degree of dyebath exhaustion as a function of time describes the rate and extent of the dyeing process. For a single dye, the exhaustion is defined as

the mass of dye taken up by the material, divided by the total initial mass of dye in the bath. For a bath of constant volume, this can be expressed by Eq. 1:

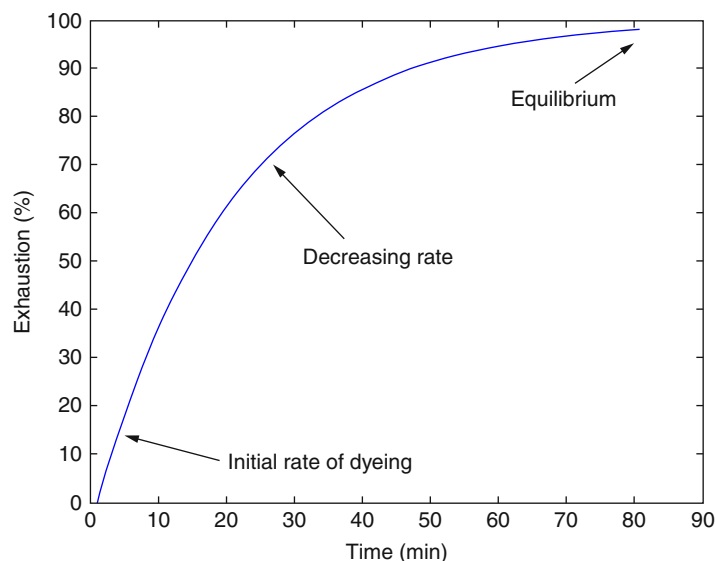
$$\% \text{Exhaustion} = (C_0 - C_t)/C_0 \quad (1)$$

where C_0 and C_t denote the concentrations of dye in the dyebath initially and at some time, t , during the process, respectively.

Exhaustion curves, such as that shown in Fig. 3, may be determined at a constant dyeing temperature, or under conditions where the temperature and other dyeing variables are changing. For many dyeings, a gradual increase of the dyeing temperature controls the rate of exhaustion, aided possibly by the addition of chemicals such as acids or salts. In cases where the dyes in the deeply dyed fibers are not able to desorb into the bath and then be redistributed onto paler fibers, such control is essential to ensure that the final color is as uniform as possible. Such redistribution of dyes is called migration.

The slope of a dyeing exhaustion curve (Fig. 3) defines the rate of dyeing at any instant during the process. The rate of dyeing gradually decreases until, if dyeing is continued long enough, an equilibrium is reached where no more dye is taken up by the fibers. There is now a balance between the rates of absorption and desorption of the dye. The equilibrium exhaustion is the maximum possible exhaustion under the given conditions. The lack of any further increase in exhaustion does not necessarily mean that a true equilibrium exists. It is possible for the dye in solution to be in equilibrium with the dye located on the outer surfaces of the fibers. True equilibrium only exists when the dye in solution is in equilibrium with the dye that has fully penetrated into the center of the fibers. Dyeings rarely continue to this point since it may take a relatively long time to attain. In fact, many commercial dyeings barely reach the point of constant exhaustion.

There are two basic methods of achieving a level exhaustion dyeing in any dye/fiber system; the first is by dye migration, and the second is by controlled dye exhaustion. The first method involves exhausting all of the dye onto the fiber and then allowing it to migrate between the fibers

Coloration, Textile,**Fig. 3** Dyebath exhaustion as a function of time [5]

in order to “level out” the dyeing. These are dyes which are able to migrate from the fiber back into the liquor and then transfer back to the fiber. This redistribution of dye improves the levelness of the dyeing and normally takes place when the dye liquor is at the boil. In this method the dye is not completely exhausted onto the substrate, and this can lead to poor reproducibility of color, and hence, additions of dye to correct the final shade are often necessary.

The second method is to ensure that the dye is exhausted in a level manner from the start of the dyeing. In this method, the dyeing rate is controlled by changing the parameters of the dye bath at a controlled rate so that the dye is deposited on the yarn in a uniform manner throughout the substrate. Careful control of these parameters, such as dyeing temperature, pH, or amount of electrolyte and flow rate, is often necessary to obtain level, well-penetrated dyeings. This is essential if the dye initially absorbed is unable to migrate from heavily dyed to poorly dyed areas during the process.

Exhaustion Profiles

Variation of the concentration of dye in the dyebath during the dyeing is referred to as the

exhaustion profile, and the shape of this profile has been believed by many researchers to be the most determining factor in levelness of dyeing.

Exhaustion control has been developed theoretically and in the laboratory by several workers. These workers used knowledge of the dyeing kinetics to devise a time/temperature profile to give a particular exhaustion profile; others have attempted a direct control of the exhaustion rate.

Studies of the theoretical basis of the relationship between levelness of dyeing and the rate of dye uptake by textile substrates were initiated in the 1950s and the 1960s. Since then there have been many investigations into the methods of controlling the exhaustion of dye bath in order to improve levelness. This has been an area of much disagreement among researchers.

Linear Exhaustion Profiles

Carbonell et al. [6] developed a mathematical representation of various exhaustion profiles and went on to calculate practical time/temperature profiles that would result in linear exhaustion profiles. Later work aimed at establishing detailed kinetic relationships in order to carry out

“isoreactive” dyeings, in which the dyeings have a linear exhaustion profile.

Cegarra et al. [7] later modified this approach to apply it to dyeings that used continuous addition (or integration) of dye into the dyebath. These dyeings were carried out at constant temperature, using a predetermined dye addition profile to achieve linear exhaustion. This method was defined as Integration Dyeing, which can be used to control the dye absorption during the integration, so as to avoid the possibility of initially fast and anomalous absorption, which may cause unlevel dyeings. In practice, this method is often used to improve the levelness, when all the dyes are added at the beginning of the process.

Several authors have stated that linear exhaustion is most likely to give a level result and developed a control strategy for automation of a dyeing machine such that the percentage exhaustion per circulation never rises above the critical value for levelness.

Other Exhaustion Profiles

The use of linear exhaustion profiles for the control of dyeing process is by no means generally accepted. A number of researchers have stated that a rapid uptake at the start of the process, with a gradual slowing of the exhaustion thereafter, should give a better result than a linear profile. The argument has been that the critical part of the exhaustion phase is the final phase, where the amount removed from the bath is large compared to that remaining, leading to a greater risk of unlevelness. Two profiles of this type, i.e., exponential and one with the exhaustion proportional to square root time, have been suggested.

Experimental work suggests that both exponential and square root profiles give a clear improvement over both a linear profile and a standard constant temperature ramp dyeing. It has been recommended to devise a reliable concentration monitoring system to control the exhaustion process. Despite the suggested strategies this is still an area requiring further research and development.

Practical Difficulties Involved in the Dyeing Process

There are some practical difficulties in achieving a quality dyeing when considering dyeing process control. These are briefly described in the following sections.

Dyeing Rate

Dyeing rates are of greater practical significance than the exhaustion at equilibrium because continuation of dyeing to equilibrium is not economical. Long dyeing times increase the risk of fiber damage and dye decomposition, particularly at higher dyeing temperatures. On the other hand, very rapid dyeings will usually result in the color being unlevel. This implies that dyeing processes should be neither too slow nor too fast. In order that dyes are used economically and as little as possible is wasted in the dyehouse effluent, the dyer prefers a high degree of exhaustion in a relatively short dyeing time. However, dyeing must be controlled so it is not so rapid that it is difficult to produce a level dyeing. If there is a strong affinity between dyes and fibers and the conditions are not controlled, a rapid strike of dyestuff will occur which will often result in unlevel dyeings. To control the rate of dye uptake, a number of dyeing parameters, including temperature, pH, electrolyte concentration, and agitation among other parameters may have to be controlled.

The slope of the exhaustion curve gives information on the rate of dyeing. The determination of these curves, however, requires much work, and they are dependent on the dyeing conditions and the nature of the goods. The dyeing rate is influenced by the temperature and by chemicals such as salts and acids, all of which also influence the final exhaustion. A clear distinction of the effects of process variables on the dyeing rate and on the final exhaustion at equilibrium is essential in the successful control of the dyeing process.

Initial Stage of Dyeing

The initial rate of dyeing (the initial slope of exhaustion versus time) is called the strike. Rapid strike by a dye often results in initial unlevelness and must be avoided for those dyes

that cannot subsequently migrate from heavily to lightly dyed areas of the fabric. For dyes of rapid strike, the dyeing conditions must limit the initial rate of exhaustion and therefore improve the levelness of the dyeing. The strike depends on the dyeing temperature, the dyeing pH, and the presence of auxiliaries.

Even for dyes of moderate and low strikes, the objective of uniform dyeing of the fiber mass is rarely achieved during the initial stages of the operation. This is because of irregularities in the material's construction, in the fiber packing, and in the distribution of residual impurities, as well as differences in temperature and flow rate of the solution in contact with the fibers.

Dye/Fiber Types and Dye Migration

Different types of textile fiber require different kinds of dyes, and in general dyes which are suitable for one type of fiber will not dye other types effectively.

The degree of exhaustion of a dye at equilibrium is higher the greater the substantivity of the dye for the fiber being dyed. Often, a very substantive dye will give a high initial rate of absorption, or strike. Substantivity is the "attraction" between dye and fiber whereby the dye is selectively absorbed by the fiber and the bath becomes less concentrated.

The ability of a dye to migrate and produce a level color, under the given dyeing conditions, is obviously an important characteristic. It can overcome any initial unlevelness resulting from a rapid strike. Migration of the dye demonstrates that the dye can be desorbed from more heavily dyed fibers and reabsorbed on more lightly dyed ones. This is especially important in package dyeing where uniform color of the yarn throughout the package is essential.

While migration is important for level dyeing, it has two major drawbacks. Firstly, dyes with greater ability to desorb from dyed fibers during migration will usually have lower fastness to washing. Dyes of very high washing fastness are essentially nonmigrating dyes for which level dyeing depends upon very careful control of the rate of dye uptake by the material. The second problem with migrating dyes is that good

migration may result in lower exhaustion, again because of their ability to desorb from the fibers.

Factors Affecting Levelness of Dyeing

As has been indicated dyeing is a very complicated process with different phenomena occurring simultaneously. Unlevelness can arise in many forms such as the unlevelness between sides, at ends of a fabric, or on the layers of a yarn package. The causes are as numerous as the effects. A quantitative analysis of the effects of many factors which influence the levelness of dyeing is a very difficult task since it would require the development of a mathematical model involving a significant number of parameters. Also, to investigate the transfer of dye through the substrate will involve the solution of nonlinear partial differential equations. Little work has been done on this aspect, according to the literature.

Flow Rate

During the dyeing process, the supply of dye through the solution to the surface of the fibers/yarns can occur in two ways, either by aqueous diffusion of dye through the liquor or by convective movements of the fluid which replace the depleted solution by fresh solution. Diffusion is a much slower process than the convective transport of dye, except at very low velocities of liquor flow.

However, an exact solution to the problem of convective diffusion to a solid surface requires first the solution of the hydrodynamic equations of motion of the fluid (the Navier-Stokes equations) for boundary conditions appropriate to the mainstream velocity of flow and the geometry of the system. This solution specifies the velocity of the fluid at any point and at any time in both tube and yarn assemblies. It is then necessary to substitute the appropriate values for the local fluid velocities in the convective diffusion equation, which must be solved for boundary conditions related to the shape of the package, the mainstream concentration of dye, and the adsorptions at the solid surface. This is a very difficult procedure even for steady flow through a yarn package of simple shape.

Measurement of Levelness

Although objective measurements have been proposed, the measurement of levelness and its causes is difficult. The levelling ability of dyes is routinely tested under a fixed set of circumstances, but the effect of changing circumstances is less often reported. Similarly, a distinction between levelness of strike and levelness from migration is not usually studied. A relatively simple means of determining levelness is via colorimetric measurement of dyed object in several locations representative of the entire substrate. A color difference metric may be used to determine the degree of variability across the substrate, and a tolerance volume may be established for this purpose. Visual assessments are also common, although these would be subjective and open to interpretation. A common procedure is to place side center side of a full width of dyed fabric to determine variations from center to sides due to variations in processing conditions. Such variability is called tailing. Another common type of variability in continuous operations is due to variability in the composition of the trough during the dyeing process that may generate shade variation between the beginning and the end of a roll, commonly denoted ending. Some online measurement systems have been proposed and used in some cases, in which digital imaging or colorimetric systems are installed in a continuous processing line which can alert the operator to “larger than expected” variations in color during the operation.

Dyes

A dye is a substance capable of imparting its color to a given substrate, such as a textile fiber. A dye must be soluble in the application medium, usually water, at some point during the coloration process. It must also exhibit some substantivity for the material being dyed and be absorbed from the aqueous solution.

For diffusion into a fiber, dyes must be present in the water in the form of individual molecules. These are often colored anions, for example, sodium salts of sulfonic acids. They may also be cations such as mauveine or neutral molecules with slight solubility in water, such as disperse dyes. The dye must have some attraction for the fiber under the dyeing conditions so that the

solution gradually becomes depleted. In dyeing terminology, the dye has substantivity for the fiber and the dyebath becomes exhausted.

The four major characteristics of dyes are:

1. Intense color
2. Solubility in water at some point during the dyeing cycle
3. Substantivity for the fiber being dyed
4. Reasonable fastness properties of the dyeing produced

The structures of dye molecules are complex in comparison with those of most common organic compounds. Despite their complexity, dye structures have a number of common features. Most dye molecules contain a number of aromatic rings, such as those of benzene or naphthalene, linked in a fully conjugated system. This means that there is a long sequence of alternating single and double bonds between the carbon and other atoms throughout most of the structure. This type of arrangement is often called the chromophore or color-donating unit. The conjugated system allows extensive delocalization of the p electrons from the double bonds and results in smaller differences in energy between the occupied and unoccupied molecular orbitals for these electrons. At least five or six conjugated double bonds are required in the molecular structure for a compound to be colored.

Table 1 gives partial classifications of dyes as presented in the Colour Index International [8]. In order to gain an optimum result, the appropriate dye class for the fiber must be used, along with specific dyeing conditions. The ten major dye classes involve acid, metal complex, mordant, direct, vat, sulfur, reactive, basic, disperse, and azoic dyes. Some of ten major dye classes shown in Table 1 can be used to dye the same fiber type, but varying conditions are required. For example, acid, metal complex, mordant, and reactive dyes can all be used to dye wool. However, there may be one type of dye that is preferred for a certain dyeing process, for example, disperse dyes for polyester fibers.

There are numerous factors involved in the selection of dyes for coloring textile materials in a particular shade. Some of these are:

Coloration, Textile, Table 1 Classification of dyes according to chemical constitution and usage

Classification of dyes according to chemical constitution	Classification of dyes according to textile usage
Azo	Direct
Anthraquinone	Azoic
Heterocyclic	Vat
Indigoid	Sulfur
Nitro	Reactive
Phthalocyanine	Acid
Polymethine	Basic (cationic)
Stilbene	Disperse
Sulfur	Mordant (metal complex)
Triarylmethane	

- The type of fibers to be dyed
- The form of the textile material and the degree of levelness required – level dyeing is less critical for loose fibers, which are subsequently blended, than it is for fabric
- The fastness properties required for any subsequent manufacturing processes and for the particular end use
- The dyeing method to be used, the overall cost, and the machinery available
- The actual color requested by the customer

The approximate relative annual consumption of the major types of fibers and dyes estimated in the year 2000 indicates that dyes used for cotton (the most widely used natural fiber) and for polyester (the most widely used synthetic fiber) dominate the market. In the case of cellulosic fibers including cotton, reactive dyes due to possessing excellent fastness properties upon fixation and demonstrating bright and brilliant shades occupy the lion's share of the market for this fiber category. Disperse dyes also occupy a large sector of the market due to their use on polyester fibers. Other colorants occupy smaller sections of the market and their applications are specific and less common.

Sorption Isotherms

In order to determine the thermodynamics of dye sorption by various fibers three main models have

been proposed. These are known as Nernst, Langmuir, and Freundlich sorption isotherms [9]. These models determine the relationship between the concentration of dye in fiber and that in solution under isothermal dyeing conditions. In the simplest form the relationship is linear (Nernst). The Nernst isotherm indicates that the distribution of dye between fibers and bath is simply due to the partition of the dye between two different solvents until one becomes saturated. In a more complex scenario, dye "sites" are present in the fiber which result in an apparent saturation value. These sorption processes are defined by the Langmuir isotherm. In some cases an empirical power function, represented by the Freundlich model, can be used to determine the relationship between the amount of dye in solution and that in fiber. An examination of this model shows that the exhaustion of dye drops towards the end of the dyeing cycle. The drop in percent exhaustion with an increase in the depth of dyeing is well known and reflects the loss of activity of the dye in solution with increasing concentration. According to this model there appears to be no saturation value for the fiber, and as more and more dye is added to the bath, more and more is taken up. There is of course a practical limit as to how much dye may be placed on the fiber. An example of a system that can be described by the Nernst isotherm is the dyeing of polyester fibers with disperse dyes. Acid dyes on wool or basic dyes on acrylic are attracted to specific dye sites with opposite charge, and these would exhibit Langmuir-type relationships. The adsorption of direct dyes on cotton may be described by a Freundlich model however, where no specific dye sites are present in fiber but a gradual decrease in the overall rate of dye adsorption is witnessed over time. In general Freundlich models are indicative of nonionic or relatively weak bonding possibilities between dye and fiber.

Summary

Many aspects of dyes and dyeing have not been covered. This is due to the complexity of the process and the large number of variables and processes involved. Textile coloration is a large

industry, and a number of excellent resources are available that cover the fundamentals of coloration and the dyeing of specific types of fibers. The reader should refer to additional resources for a detailed examination of the topics.

Cross-References

- [Coloration, Fastness](#)
- [Coloration, Mordants](#)
- [Colorant, Natural](#)
- [Colorant, Textile](#)
- [Colorant, Environmental Aspects](#)
- [Dye](#)

References

1. Vickerstaff, T.: *The Physical Chemistry of Dyeing*, 2nd edn. Oliver and Boyd, Edinburgh (1954)
2. Morton, W.E., Hearle, J.W.S.: *Physical Properties of Textile Fibres*, 3rd edn. Textile Institute, Manchester (1993)
3. Shamey, R., Zhao, X.: *Modelling, Simulation and Control of the Dyeing Process*. Woodhead Publishing, Cambridge, The Textile Institute (2014)
4. Johnson, A.: *The Theory of Coloration of Textiles*. Textile Institute, Bradford (1989)
5. Zhao, X.: *Modelling of the Mass Transfer and Fluid Flow in Package Dyeing Machines*, Ph.D. Thesis, Heriot-Watt University (2004)
6. Society of Dyers and Colourists: *Colour Index International*. Society of Dyers and Colourists, Bradford (1989)
7. Carbonell, J., Knobel, R., Hasler, R., and Walliser, R.: Mathematische Erfassung der Zusammenhänge zwischen Farbekinetik und Flottdurchfluss in bezug auf die Homogenität der Farbstoffverteilung auf der Faser in Systemen mit Flotenzirkulation. *Melliand Textilber.* **54**, 68–77 (1973)
8. Cegarra, J., Puente, P., Valldeperas, J., Pepio, M.: Dyeing by Integration. *Text Res J.* **58**, 645–653 (1988)
9. Aspland, R.: *Textile Dyeing and Coloration*. American Association of Textile Chemists and Colorists (1997) Research Triangle Park, NC, USA

Colorfulness Adaptation

- [Color Contrast Adaptation](#)

Coloring

- [Coloration, Textile](#)

Color-Opponency, Unique Hues

Sophie Wuerger¹ and Eriko Self²

¹Department of Psychology, University of Liverpool, Liverpool, UK

²Department of Psychology, California State University, Fullerton, CA, USA

Synonyms

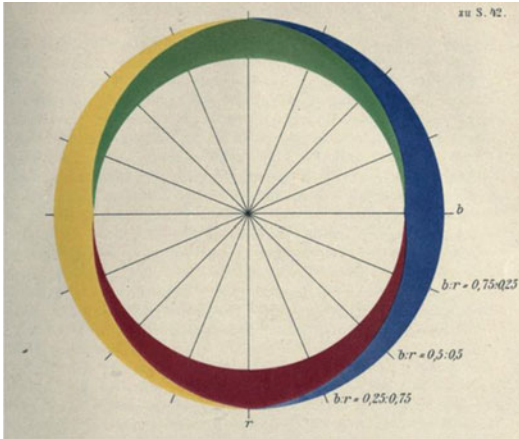
[Color-opponent processing](#); [Focal colors](#); [Hue opponency](#)

Definition

Unique hues are colors that cannot be described by any other color term than their own. There are four unique hues: red, green, blue, and yellow. Unique hues emerge at the equilibrium points of the color-opponent mechanisms: a color which appears neither reddish nor greenish is a unique yellow or unique blue; a color which appears neither yellowish nor blueish is unique red or unique green.

Historical View

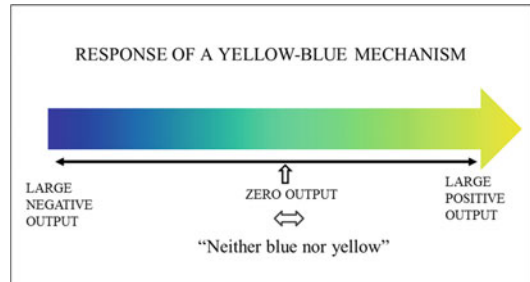
Four colors occupy a special place in color perception. They are called unique hues and they originate from the opponent color theory proposed by Ewald Hering in 1878 [1]. Hering was the first to notice that some pairs of colors, namely, red and green, and yellow and blue, cannot be perceived at the same time. He named these pairs of colors “Gegenfarben” [opponent colors], since they are mutually exclusive colors as indicated by the lack of overlap between red and



Color-Opponency, Unique Hues, Fig. 1 Hering [1], page 42. The opponent colors, red vs. green, and yellow vs. blue, are mutually exclusive. This idea is conveyed in Hering's original figure by showing that there is no overlap between red and green, or between yellow and blue

green, and between yellow and blue in Hering's original figure (Fig. 1). Hering's color vision theory postulates three opponent processes: two chromatic processes of red-green and blue-yellow and one achromatic process of light vs. dark. Unique hues are perceived when one of the two chromatic processes is polarized in one direction and the other opponent process is at equilibrium. For example, the putative yellow-blue mechanism signals colors from yellow to blue; if the stimulus contains neither yellow nor blue, then it elicits no response (Fig. 2) and this mechanism is deemed to be at equilibrium. Stimuli that are perceived as "neither yellow nor blue" are defined as "unique red" or "unique green," depending on which side of the neutral grey point they are located. Likewise, colors that silence the red-green opponent mechanism are called "unique yellow" or "unique blue". Hurvich & Jameson [1] were the first to use a hue cancellation procedure to find the equilibrium points (null responses) of Hering's color-opponent mechanisms.

The opponent-color vision theory was initially thought to be at odds with the trichromatic theory proposed by Thomas Young in 1802 and developed by Hermann von Helmholtz in 1850. The trichromatic theory states that there are three types of receptors responsible for conveying color

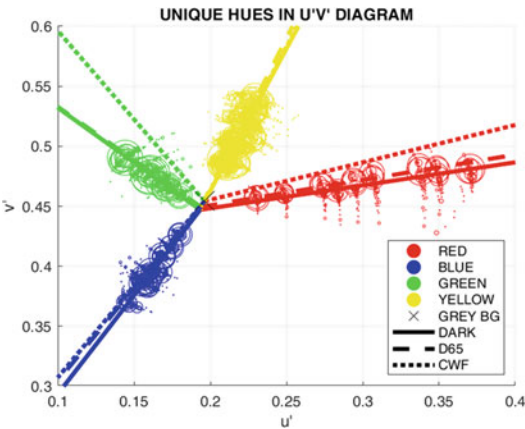


Color-Opponency, Unique Hues, Fig. 2 The response of a hypothetical yellow-blue mechanism is shown

signals. These two theories of color vision were the focus of controversy among scientists until they were integrated into the two-stage model in the 1960s [2]. The trichromatic theory holds at the receptor level in the retina where there are three types of cones. These are called short-, medium-, or long-wavelength-sensitive (S, M, or L) cone receptors, and the names are derived from the different spectral region of the peak in their spectral sensitivities. After light is absorbed in the three classes of cone receptors (S, M, L), the outputs of the receptors are combined in an opponent fashion, both at the retinal ganglion cell level and again at later processing stages in the visual cortex [3]. Therefore, the trichromatic theory of color vision and the opponent-color theory are not at odds with each other; rather they reflect different processing stages in the visual pathway, namely, photoreceptors in the retina versus retinal ganglion cells/LGN and cortical processing.

Behavioral Evidence for Unique Hues

Color-normal observers are readily able to adjust lights such that they appear "neither yellowish nor blueish" or "neither reddish nor greenish" without any training or detailed instruction, but attempts to provide conclusive behavioral evidence for the privileged role of these four hues have been unsuccessful. Traditionally, unique hues have been measured with spectral lights created by a monochromator to show a light of only one wavelength [4]. More recently, Kuehni and colleagues [5] have used broadband stimuli such as those



Color-Opponency, Unique Hues, Fig. 3 Unique hue settings for 185 color-normal observers [11] are plotted in an approximately uniform $u'v'$ diagram. Symbols are proportional to the number of data points. The solid, dashed, and dotted lines indicate the first principal component for the dark, D65 (daylight simulator) and CWF (office light) viewing conditions, respectively

printed on paper (Munsell Color chips; Natural Colour System chips [6]), while stimuli presented on a computer screen have also been widely used [7].

Larimer and colleagues [8, 9] were the first to test the linearity and additivity of the color-opponent mechanisms. The red-green equilibria are approximately linear; that is, unique yellow and blue are co-linear; the yellow-blue equilibria on the other hand are not linear, that is, unique red and green do not lie on a line through the origin, the neutral grey background (Fig. 3; [10]). Furthermore, unique yellow and unique blue are fairly color constant in comparison to unique green and red: the unique yellow and blue settings are not affected by the changes in the ambient illumination (“Dark,” “D65,” “CWF,” Fig. 3).

Unique hues are defined as the hues that emerge at the equilibrium points of the putative color-opponent mechanisms [8, 9] (see Fig. 2). In the two-dimensional $u'v'$ chromaticity diagram, this can be formalized as follows:

$$YB_R = \alpha_R \times \Delta u' + \beta_R \times \Delta v' = 0 \quad (1)$$

where $\Delta u'$ and $\Delta v'$ are the incremental chromaticity coordinates of the unique red settings with

Color-Opponency, Unique Hues, Table 1 Coefficients of mechanisms at equilibrium

Unique hue	Orthogonal vector	Orthogonal vector coefficients (SD)	
		α	β
Red	YB_R	$-0.206 (0.050)$	$0.977 (0.008)$
Green	YB_G	$0.408 (0.511)$	$0.526 (0.545)$
Yellow	RG_Y	$0.926 (0.025)$	$-0.370 (0.063)$
Blue	RG_B	$0.838 (0.028)$	$-0.542 (0.049)$

respect to a grey background. α_R and β_R are the respective weights of the yellow-blue mechanism (YB_R) which is silenced by unique red. Similarly, the orthogonal vector for unique green is defined as:

$$YB_G = \alpha_G \times \Delta u' + \beta_G \times \Delta v' = 0 \quad (2)$$

To generate unique yellow (unique blue), a red-green opponent mechanism (RG) is assumed to be at equilibrium, hence the orthogonal vectors are defined as follows:

$$RG_Y = \alpha_Y \times \Delta u' + \beta_Y \times \Delta v' = 0 \quad (3)$$

$$RG_B = \alpha_B \times \Delta u' + \beta_B \times \Delta v' = 0 \quad (4)$$

To a first approximation, the unique green vector is orthogonal to the unique blue and the unique yellow lines (cf Fig. 3), hence the direction of the RG_{YB} mechanism is aligned with the unique green line (Table 1). This means that observers use a hue close to unique green when asked to cancel the greenish tint in a hue cancellation task. In contrast, the unique red line is not aligned with the orthogonal vector for unique yellow or blue ($RG_{Y,B}$). Hence, when observers are asked to cancel the reddish tint in a hue cancellation task, their “internal red” is not the same as their unique red settings, but is much more bluish than their unique red.

The lack of collinearity of the unique red-green line and the failure of orthogonality of these two hue-opponent mechanisms is not a consequence of the particular chromaticity diagram used, but is an intrinsic property of hue opponency which is

also revealed in the a^*b^* chromaticity diagram (CIECAM02) and physiologically motivated chromaticity diagrams (see Fig. 4 below).

Neural Basis of Unique Hues

The discovery of neurons with chromatically opponent receptive fields in the retina and the Lateral Geniculate Nucleus (LGN) [11] initially led to the erroneous suggestion that these neurons constitute the physiological substrate of Hering's color-opponent processes and consequently, the unique hues. It is now clear that these early cone-opponent mechanisms are not the neural substrate of the behaviorally measured hue opponency [3, 12]. The discrepancy between hue opponency and cone opponency is readily seen by re-plotting the unique hue loci in a cone-opponent diagram (Fig. 4) where the x axis reflects the activity of the L-M cone-opponent mechanism and the y axis that of the S-(L + M) cone-opponent mechanism. These cone-opponent color directions are sometimes referred to as "cardinal directions" of color space [12].

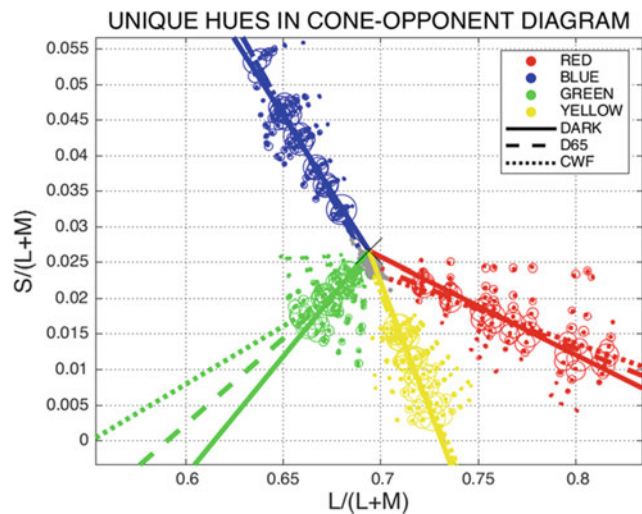
The unique hue lines are not aligned with the cone-opponent axes [12]. The red-green equilibria (=line connecting unique yellow and blue) are not parallel to the S/(L + M) cone axis; the yellow-

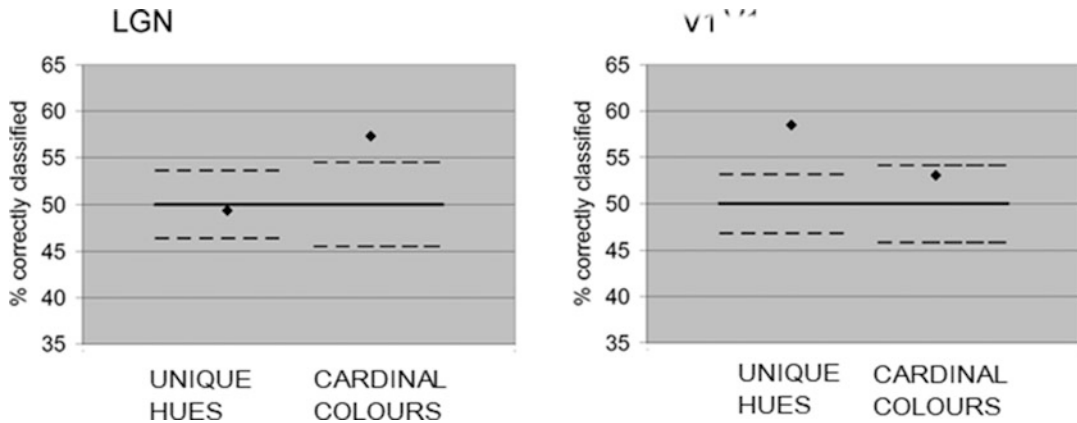
blue equilibria (=line connecting unique red and green) are not aligned with the L/(L-M) axis. Unique blue requires, in addition to the S cone input, an M cone input; unique yellow requires a negative S cone input, and a positive L cone input. Similarly, unique green requires a negative S cone and a positive M cone input. Unique red is the only hue that is at least approximately aligned with the L/(L + M) axis; but even here a systematic negative S cone input is required. Since early cone-opponent mechanisms do not constitute the neural mechanisms underlying the observed hue opponency and the resulting unique hues, further re-combinations of the early cone-opponent signals must occur between the LGN and the primary visual cortex, or within the visual cortical areas [3, 13].

The cone-opponent signals from the LGN undergo another chromatic transformation when reaching the primary visual cortex [13]. This re-combination of color signals was directly tested using fMRI decoding techniques in the LGN and visual cortex [14, 15]. Modulations along the cone-opponent (red-green, lime-violet axes in Fig. 4; denoted as "Cardinal Colors" in Fig. 5) and light-dark directions can be decoded in LGN, but not in the primary visual cortex, V1. Conversely, unipolar modulations between the grey background and a particular unique hue can be

Color-Opponency, Unique Hues,

Fig. 4 Unique hue settings for 185 color-normal observers [11] are plotted in the cone-opponent diagram based on the CIE 2006 2-deg cone fundamentals. The grey background is shown as an 'x'. Lines denote first principal components





Color-Opponency, Unique Hues, Fig. 5 Classification performance for cardinal colors and for unique hues in the LGN (left) and primary visual cortex (right). Dashed lines indicate 95% confidence limits

decoded in V1, but not in LGN (denoted as ‘Unique Hues’ in Fig. 5).

Successful classification of cardinal colors in the LGN is consistent with the known spatial clustering in layers of neurons tuned to cone-opponent color modulations [11] and the failure to decode cone-opponent modulations in V1 is consistent with cortical chromatic tuning properties different from those found in the LGN. The successful decoding of unipolar hue modulations corroborates evidence of a hue map in V1, but does not constitute strong evidence for the special status of unique hues.

Individual Differences

To a first approximation, unique hue judgments are robust across different cultures and languages [16], but there are variations in the unique hue loci among people with normal color vision [17]. The sources of these individual differences remain unknown. Possible physiological sources of individual differences include pre-receptoral light filtering by the lens and macular pigment, spectral sensitivity and optical density of cones, relative number of different types of cones, and relative sensitivity of the two chromatic mechanisms at the second stage. Among these possibilities, variations in pre-receptoral filtering and optical

density of cones are not large enough to account for observed individual differences in unique hues [18]. The L/M cone ratio can vary by more than a 30-fold range among individuals, and it is far too large to account for the range of observed unique hue differences [17] and unique yellow settings have been shown to be independent of the L:M cone ratio [19]. Furthermore, no robust relationship has been found between the variability in the cone-opponent mechanisms and the variability in unique hue settings, hue scaling, and hue naming tasks [20]. In contrast to chromatic sensitivity, unique hue loci are remarkably stable across the lifespan [10, 21].

Cultural and Environmental Factors

Since unique hue loci are not significantly constrained by optical factors or low-level color mechanisms, it is likely that environmental factors play a role in the emergence of unique hues. For example, unique hue loci correspond to surface reflectance that are optimally color constant when illuminated by natural daylight [22]. However, cultural factors are entwined with environmental factors since the distribution of colors available in a particular environment will shape the range of colors utilized by a particular culture. Differences in unique hue loci between observers in India and

the United States have been documented [23], but the inter-group variations are relatively small compared to the within-group variation. Differences in unique hue loci across different cultures are likely to reflect the interplay between particular languages spoken, cultural norms, and environmental factors. The neural plasticity of these higher-order color mechanisms may be useful to compensate for age-related physiological changes such as the yellowing of the lens, differences in the retinal make-up including the L:M cone ratio, and changes in the chromatic environment [24].

Unique Hues and Focal Colors

To evaluate the universality of color terms across languages, color naming data from native speakers of 20 languages were collected [25]. A palette of Munsell color chips was presented to human observers with the instruction to choose a best example of the color term and to draw a boundary of the chips that can be called by the color term. Though their origins are completely different, unique hues and focal colors for categories red, green, blue, and yellow have been shown to closely correspond with each other empirically [26].

Summary

Unique hues are the four primary colors in color perception. The neural basis of the four unique hues is still a mystery, but it is clear that the cone-opponent mechanisms in the early visual pathway are not the neural substrates of the unique hues. These early cone-opponent mechanisms must undergo further transformations to lead to a representation of the unique hues in cortical visual areas.

Unique hues are robust across the life span but individual differences in the unique hue loci exist. The sources of these individual differences have not yet been identified, but peripheral

physiological factors have been excluded as the main source of individual variation. The emerging picture is that unique hues are shaped by both environmental factors resulting in long-term adaptation as well as cultural factors.

Cross-References

- [Berlin and Kay Theory](#)
- [Environmental Influences on Color Vision](#)
- [Focal Colors](#)
- [Primary Colors](#)
- [Psychologically Pure Colors](#)

References

1. Hurvich, L., Jameson, D.: An opponent-process theory of color vision. *Psychol. Rev.* **64**, Part 1(6), 384–404 (1957)
2. Mollon, G., John, D., Jordan: In: Dickinson, C., Murray, I., Carden, D. (eds.) *On the nature of unique hues*. Taylor & Francis (1997)
3. De Valois, R.L., De Valois, K.K.: A multi-stage color model. *Vis. Res.* **33**(8), 1053–1065 (1993)
4. Volbrecht, V., Nerger, J., Baker, L., Trujillo, A., Youngpeter, K.: Unique hue loci differ with methodology. *Ophthalmic Physiol. Opt.* **30**, 545–552 (2010)
5. Hinks, D., Cárdenas, L.M., Kuehni, R.G., Shamey, R.: Unique-hue stimulus selection using Munsell color chips. *J. Opt. Soc. Am. A.* **24**(10), 3371–3378 (Oct. 2007)
6. Shamey, R., Zubair, M., Cheema, H.: Effect of field view size and lighting on unique-hue selection using Natural Color System object colors. *Vision Res.* **113**(PA), 22–32 (2015)
7. Webster, M.A., Miyahara, E., Malkoc, G., Raker, V.E.: Variations in normal color vision II Unique hues. *J. Opt. Soc. Am. A.* **17**(9), 1545 (2000)
8. Larimer, J., Krantz, D.H., Cicerone, C.M.: Opponent process additivity-II. Yellow/blue equilibria and non-linear models. *Vis. Res.* **15**(6), 723–731 (1975)
9. Larimer, J., Krantz, D.H., Cicerone, C.M.: Opponent-process additivity-I: red/green equilibria. *Vis. Res.* **14**(11), 1127–1140 (1974)
10. Wuerger, S.: Colour constancy across the life span: Evidence for compensatory mechanisms. *PLoS One.* (2013)
11. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (Dec. 1984)

12. Krauskopf, J., Williams, D.R., Heeley, D.W.: Cardinal directions of color space. *Vis. Res.* **22**(9), 1123–1131 (1982)
13. De Valois, R.L., Cottaris, N.P., Elfar, S.D., Mahon, L.E., Wilson, J.A.: Some transformations of color information from lateral geniculate nucleus to striate cortex. *Proc. Natl. Acad. Sci. U.S.A.* **97**(9), 4997–5002 (2000)
14. Wuerger, S.M., Parkes, L.: Unique hues: Perception and brain imaging. In: *New Directions in Colour Studies*, pp. 445–455 (2011)
15. Parkes, L.M., Marsman, J.-B.C., Oxley, D.C., Goulermas, J.Y., Wuerger, S.M.: Multivoxel fMRI analysis of color tuning in human primary visual cortex. *J. Vis.* **9**(1) (2009)
16. Saunders, B.A., van Brakel, J.: Are there nontrivial constraints on colour categorization? *Behav. Brain Sci.* **20**(2), 167–228 (1997)
17. Webster, M.A., Kay, P.: Individual and population differences in focal colors. *Anthropol. Color.* (2007)
18. Pokorny, J., Smith, V.C., Wesner, M.F.: Variability in cone populations and implications. *From Pigment. Percept.*, 23–34 (1991)
19. Brainard, D.H., et al.: Functional consequences of the relative numbers of L and M cones. *J. Opt. Soc. Am. A.* **17**(3), 607–614 (2000)
20. Emery, K.J., Volbrecht, V.J., Peterzell, D.H., Webster, M.A.: Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vis. Res.* **141**, 66–75 (2017)
21. Scheffrin, B.E., Werner, J.S.: Loci of spectral unique hues throughout the life span. *J. Opt. Soc. Am. A.* **7**(2), 305 (1990)
22. Philipona, D.L., O'Regan, J.K.: Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Vis. Neurosci.* **23**(3–4), 331–339 (2006)
23. Webster, M.A., et al.: Variations in normal color vision III Unique hues in Indian and United States observers. *J. Opt. Soc. Am. A.* **19**(10), 2002 (1951)
24. Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., Williams, D.R.: Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron.* **35**(4), 783–792 (2002)
25. Berlin, P., Kay, B.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
26. Kay, P., Mcdaniel, C.K.: Linguistic Society of America The Linguistic significance of the meanings of basic color terms. *Language.* **54**(3), 610–646 (2011)

Color-Opponent Processing

► Color-Opponency, Unique Hues

Colors: Where Are They?

Susan P. Farnand¹ and Rolf G. Kuehni²

¹Program of Color Science, Rochester Institute of Technology, Rochester, NY, USA

²Color Science Laboratory, Carolina State University, Raleigh, NC, USA

Statement of the Question

Most people with normal color vision would agree that ripe strawberries are red, summer leaves are green, and the sky is blue. But, are red and green actual properties of berries and leaves? Are strawberries actually red and leaves green? Or, do these colors only exist in the minds of viewers? Expert opinions, recorded first in the early seventeenth century, generally indicate a greater likelihood of the subjective nature of the color experience, comparable to other sensory experiences, located only in the minds of creatures with a color vision system. Color scientists would contend that colors are not just related to object properties but are a product of the light interacting with the object to trigger a response in the eyes and brain of the observer. To understand the subjective or objective nature of color, we must understand the properties of light and the visual system of the observer, as well as properties of the object.

The idea that colors (and other sensory experiences) exist only in consciousness was probably expressed for the first time in writing in a book published in 1623 by Galileo Galilei who said “I think that tastes, odors, colors, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness.” [1]

More recently Steven Pinker stated the following: “Our minds evolved by natural selection to solve problems that were life-and-death matters to our ancestors, not to commune with correctness.” [2] This comment makes clear that the question of the source of color perception is not unique

but that the same situation applies to all sensory perceptions, including smell, taste, and hearing.

Light as a Primary Cause of Color Perception

Color perception requires light. So, the first question to be considered is: what is light? People have argued about this for millennia, and a scientifically solid answer still does not exist. Over the centuries, light has alternatively been described as consisting of waves or particles. The weight of Isaac Newton's opinion helped bolster the argument of light as a particle until the time of Young's double slit experiment, which swayed description of light back to waves. This was reinforced by Maxwell's introduction of a theory of electromagnetic radiation (EMR) [3]. This explanation held until Einstein's Quantum Theory of Light, which described light as packets of energy known as photons. Most expert authors now describe light as having both a wave and a particle nature. In either case, it can be said that the experience of natural light is the perceptual response of most living creatures to a narrow band of electromagnetic energy that travels from the sun directly and indirectly to the earth. As wave energy, the band of wavelengths representing light (for humans) ranges from about 370 nanometers (nm) to about 750 nm. It is a small range of EMR, but enough to be able to provide considerable information about its immediate source without containing the shorter wavelengths energy that is damaging to the visual systems of most living creatures and is largely filtered by the Earth's atmosphere.

Humans have wondered for centuries if EMR in the form of perceived light from the sun also carries colors in it. An answer, based on extensive experimental experience, was provided by Newton in his 1704 book *Opticks*, in which he said "For rays to be spoken properly are not coloured. In them is nothing else than a certain power and disposition to stir up a sensation of this or that colour" [4]. This can be demonstrated, in a typical office or classroom setting, by passing a spectrally narrow beam of light, such as that from

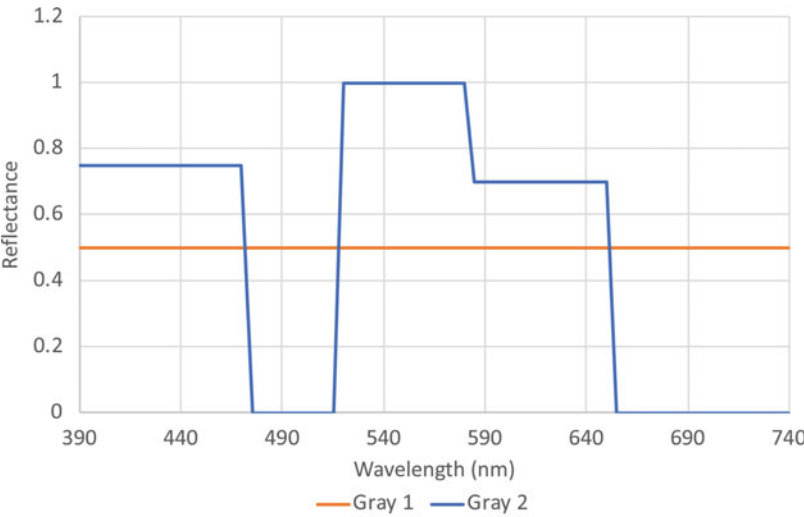
a laser pointer, approximately horizontally in front of our eyes. Unless there is dust in the air that reflects some of the light, we will not see the beam or experience any colors unless it is reflected from a surface into our eyes.

The Visual System Is the Creator of Color Perceptions

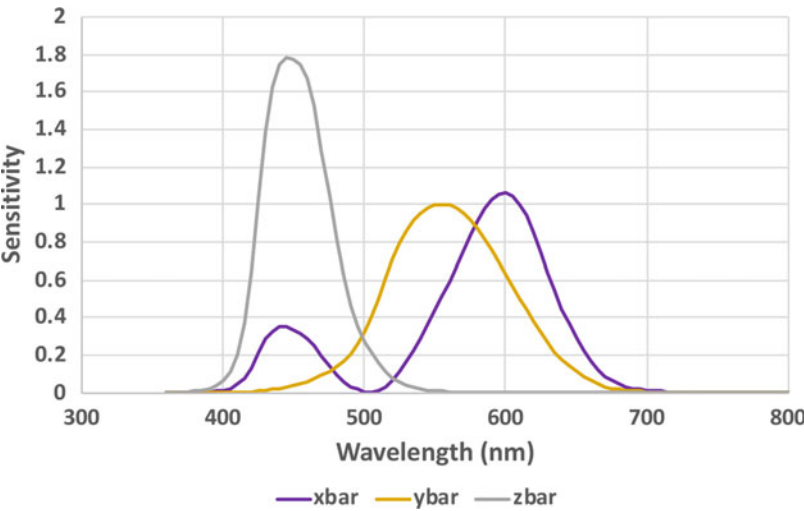
The human visual system integrates the spectral response to light and colorant mixtures. Because of this, identical perceptual results for color-normal observers to a given mixture of lights of varying wavelengths or colorants of varying reflectance functions can be obtained with many different mixtures of lights or colorants. An extreme example of this phenomenon, which is known as metamerism, is shown in Fig. 1. Two versions of reflectance functions of a medium lightness gray are depicted. When viewed by color-normal observers under standard conditions of lighting and surround, the appearance of the two samples will be identical despite the large difference in reflectance functions. This can be mathematically predicted using the system of experimentally determined color matching functions, $\bar{x}$, $\bar{y}$, and $\bar{z}$ (Fig. 2) [5]. With them, related tristimulus values X , Y , and Z can be calculated based on reflectance functions representing the color samples involved. For the reflectance functions of Fig. 1 and a uniform light source under which the samples are viewed, the tristimulus values X , Y , and Z are identical for both samples, indicating that they will be perceived by color-normal humans as identical in color. A large number of samples having the same tristimulus values but different reflectance functions can be created, depending on the available colorants.

In the metamerism example described above, the two sample colors having different spectral curves that match under one light source may appear to have quite different colors under another light source. Because the visual response integrates the spectral curves of the light source and the sample relative to the sensitivity of the viewer (the personal color matching functions), observer metamerism may also occur. In this case, two

Colors: Where Are They?,
Fig. 1 Two theoretical reflectance functions representing colored materials that by the CIE 1931 standard observer are seen as having identical gray appearance, based on the related color matching functions of Fig. 2. In both cases X, Y, and Z have identical values of 50.0 each



Colors: Where Are They?,
Fig. 2 CIE 1931 standard colorimetric, 2°, color matching functions. The colors used are arbitrary and do not represent any relation to perceived colors



spectrally different colorants or lights may match for one color-normal observer but less so or not for another, due to small but sometimes significant differences in human visual systems. Metamerism means that the trichromatic human visual system is seriously limited in its capability of distinguishing between different sets of EMR arriving at the eyes as reflected from the surface of the colored samples or from self-luminous stimuli.

As mentioned, a perceptual response to light requires a system that is sensitive to EMR and is capable of interpreting in some way the responses of the sensors. Without this, we would not be

aware of lights, as fully blind people can confirm. The animal/human experience of light is the result of the specific response of the visual system to the stimulation by selected EMR levels entering the eyes. A system with a single type of sensor (a monochromatic system) cannot provide distinction between spectrally different levels of energy (or wavelengths) of the radiation impacting on it. At least two or more sensor types with sensitivities to different wavelength ranges are needed, together with an interpretation system, to provide a degree of distinctiveness of different wavelength levels of the EMR.

The development of such a system began with light-sensitive cells, which were originally likely used for orientation of the organism toward or away from the sun. All multi-cellular organisms possess the genetic toolkit for such photoreceptors. Fossil evidence indicates that the first eyes appeared early in the Cambrian period (543 to 538 MYA). At that time, “an explosion of life” occurred that resulted in the development of all the current existing phyla and that some authors, such as Andrew Parker in his book *In the Blink of an Eye*, attribute to the advent of vision [6]. Most argue that this macroevolution was more complex than that and involved such conditions as the presence of oxygen in air and a warming climate. During the Cambrian explosion, trilobites (and other living creatures) evolved which had complex seeing capabilities. They had a frontal head with two-eye structures comprising some 150 visual cells, each with its own minute lens and a single light sensor below it. The results from these sensors fed into the brain. Animals having compound eyes, including many arthropods, are still in existence today.

For the creatures with early eyes based on single kinds of light receptors, the resulting achromatic percepts of potential food, enemies, or partners were clearly very useful, but additional more detailed information regarding the nature of objects in front of their eyes was obviously also supportive. Accordingly, a kind of light sensor known as cones also began to develop. Cone sensors with sensitivity to different ranges of EMR of light resulted in more specific information about the surrounding world. This more detailed input from the cone sensors having different spectral sensitivities required more complex brain systems to interpret. When cone sensors first developed is not known, but by ca. 500 MYA, jawless vertebrate fish are believed to have had four types of cones with different spectral sensitivities [7]. The presence of two or more cone types differing in spectral sensitivity in an eye resulted in some loss of image sharpness but in a gain of more qualitative information about the fishes’ surroundings, expressed in form of color vision.

The development of mammals began around 150–200 MYA. They were mostly nocturnal and thereby not in need of four cone types. They developed simple dichromatic color vision systems. Dichromatic mammals continue to be the norm, with some important exceptions. Early primates, such as lemurs, which essentially lived in trees, required quick and detailed distinction of elements of their arboreal environment and, consequently, developed a trichromatic vision system with stereopsis, the overlapping in the brain region of information from both eyes, allowing for depth perception which was important for more accurately assessing distances of food sources or enemies. The old-world primates, including gorillas and chimpanzees, also had and have a trichromatic stereoptic visual system. Our direct ancestor, homo sapiens, began life ca 300 TYA in Africa. Like earlier human ancestral groups, they were trichromats, with short, medium and long (S, M and L) cone types in the eyes. The human M cones have spectral sensitivity relatively close to that of L cones, as Fig. 2 shows.

The electric signals generated when cone cells absorb light are transmitted to the brain through horizontal, bipolar, amacrine, and ganglion cells, all of which are located in the retina. The connections between different layers are complex. The information produced by the ganglion cells is passed through the lateral geniculate nucleus into the occipital lobe of the brain at the back of the head where additional visual processing takes place. The color perception-generating apparatus in the retina and the brain/mind is quite complex. While significant progress has been made in understanding the visual system in humans, all its details are not yet known. In whatever way it may be accomplished, the final result is the experience in our mind of colored images of the outer world. Specific colors experienced in this fashion can be related to specific wavelengths of light, but most sensations are based on broadband mixtures of EMR stimuli.

There are considerable differences in the visual systems in the animal kingdom. Creatures with color vision have between 2 and 40 different visual sensor types, resulting in different

approaches to interpreting EMR signals. The silver spinyfin, a deep-sea fish, has 2 cones and 38 rod opsins to provide detailed information in the spectral range of bioluminescence if present as well as the daylight that reaches their environment [7]. Humans with normal color vision are trichromats, having sensors with peak sensitivities in the eye at approximately 440, 530, and 558 nm.

Colors we experience are the result of the interpretation of input sensory signals in our brains/minds, as are all other sensory experiences. Metamerism demonstrates that the visual system, depending on the number of sensor types, has stronger or weaker capability to identify and distinguish between spectrally different kinds of EMR in the visible range. Different animals have different visual systems and so can distinguish different elements in their environments. For example, bees and some birds have sensitivity in the ultraviolet region, but not in the longer wavelength region of the human visible spectrum. Some insects and the mantis shrimp can detect polarized light. The visual systems of these animals are adapted to interpret environmental information according to their specific needs. There is no reason to expect that how they see nature would match how humans see it.

There is also no reason to expect that the color experience of humans with normal color vision would be fully identical. Evidence suggests that it is not. As mentioned with regard to observer metamerism, there are small but measurable differences in observer color matching functions. Understanding observer differences has been an area of active research in recent years [8, 9].

Along with metamerism, there are other phenomena that demonstrate the variable relationship between physical measurements of the interaction of objects and light and the appearance of color. These include simultaneous contrast, chromatic adaptation, and memory colors. Please refer to the relevant entries for additional information.

Light and colors remain personal experiences that we cannot describe in a meaningful objective way, as it is the case with all other sensory experiences we have. We can identify the physical or chemical stimuli that produce these experiences but not the experiences themselves. Given the

physical results of stimulus mixture, as shown in the CIE colorimetric system, we can predict the results of such stimulus mixtures but with prediction of the perceived results only in form of assumed identity of perceived color experiences for all “color-normal” observers. Experimental assessments of the stimuli of so-called primary hues red, blue, yellow, and green can result in reasonable mean agreement between groups of observers, but there is significant interobserver variability in the selections, indicating sources of variability in the perception processes.

It is an undisputable fact that the color experiences humans have are generated in our brains/minds. Are they perceptual interpretations in brain/mind of the EMR signals arriving at the eyes, based on a mechanism that life generated over millions of years? Or are they reconstructions of colors out in the world based on the EMR patterns arriving at the eyes? A scientifically based direct or indirect solid argument for colors being in the outside world has as yet not been offered.

Authors of existing scientific books on color vision generally did not address this issue. An exception is the psychologist and academic Stephen E. Palmer, author of the 1999 book *Vision Science, Photons to Phenomenology*, in which he stated (Chap. 3): “People universally believe that objects *look* colored because they *are* colored . . . The sky looks blue because it *is* blue, grass looks green because it *is* green, and blood looks red because it *is* red. As surprising as it may seem, these beliefs are fundamentally mistaken.” [10]

In *The Case Against Reality: Why Evolution Hid the Truth from Our Eyes*, [11] the author, Donald Hoffman, a professor of cognitive science, addresses in significant detail the related issues of our sensory experiences. He shows, with logically solid arguments, that the world we experience is not the real, physical world but one that evolution generated for all living systems to provide information to the system owners. It provides not physically truthful information but information that simplifies their worlds and their actions in them: colors, smells, tastes, sounds, etc.

Hoffman cannot, as yet, offer totally indisputable evidence for his claims. Presently, we still lack full understanding of the relationship between the physical world and the perceptual world of living creatures. There is much work to be done in the coming decades.

The idea that colors and other sensory experiences are in reality the interpretations of available physical information needed for survival and exist only in the bodies/minds of living creatures will likely become more and more acceptable as time goes on.

Cross-References

- [CIE Chromatic Adaptation: Comparison of von Kries, CIELAB, CMCCAT97, and CAT02](#)
- [Memory Color](#)
- [Simultaneous Color Contrast](#)

References

1. Galilei, G.: *Il sagggiatore*, p. 274. Lincean Academy, Rome (1623). (The assayer, abridged english text PDF available at Stanford University)
2. Pinker, S.: *How the Mind Works*. W. W. Norton, New York (1997)
3. Maxwell, J.C.: A dynamical theory of the electromagnetic field. *Philos. Trans. R. Soc. Lond.* **155**, 459–512 (1865)
4. Newton, I.: *Opticks*, Experiment VI, Definition, p. 90/91. Smith & Walford, London (1704)
5. Wyszecki, G., Stiles, W.S.: CIE color matching functions. In: Wyszecki, G., Stiles, W.S. (eds.) *Color Science*, 2nd edn, pp. 806–816. Wiley, New York (1982)
6. Parker, A.: *In the Blink of an Eye*. Basic Books, New York (2003)
7. Pennisi, E.: Jepper, creatures, where'd you get those peepers? *Science*. **364**(6440), 520 (2019)
8. Asano, Y., Fairchild, M.D., Blondé, L.: Observer variability experiment using a four-primary display and its relationship with physiological factors. In: *Color and Imaging Conference*, vol. 2013, no. 1, pp. 171–176. Society for Imaging Science and Technology (2013)
9. Sarkar, A., Blondé, L.: Colourimetric observer categories and their applications in colour and vision sciences. In: *Proceedings of the CIE Centenary Conference*, OP 52, pp. 393–401. Available on CIE webshop as CIE x038:2013, (2013, April)
10. Palmer, S.E.: *Vision Science, Photons to Phenomenology*. MIT Press, Cambridge, MA (1999)
11. Hoffman, D.: *The Case against Reality; why Evolution Hid the Truth from our Eyes*. W. W. Norton Co., New York (2019)

Color-Separation Overlay

- [Compositing and Chroma Keying](#)

Color-Word Interference

- [Stroop Effect](#)

Colour Index

Geoffrey Hallas

Department of Color Chemistry, University of Leeds, Leeds, UK

Definition

Online database of colorants (dyes and pigments), published jointly by the Society of Dyers and Colorists (SDC) and the American Association of Textile Chemists and Colorists (AATCC). The Colour Index uses a dual classification system in which the prime descriptor is the Colour Index Generic Name (often abbreviated to CIGN), which is related to application (usage). The other descriptor is the Colour Index Constitution Number (often abbreviated to CICN), which is related to chemical structure.

Overview

Historical

First Edition

Early attempts to list and classify synthetic dyes and pigments, involving chemical structure and hue, appeared before the end of the nineteenth

century [1]. The most successful system at this time was the Farbstofftabellen of Gustav Schultz (with seven editions between 1888 and 1932, followed by Supplements to the Seventh Edition in 1934 and 1939); each colorant was given a classification number. Meanwhile, from 1922, the Society of Dyers and Colorists published a series of 12 monthly compilations, which culminated in the First Edition of the Colour Index, edited by F. M. Rowe, as a bound volume in 1924; as well as synthetic organic colorants (1230 entries), the 750-page volume contained sections on natural organic colorants and inorganic colorants. A Supplement was published in 1928, the bulk of which consisted of a new alphabetical index of commercial names.

Second Edition

After the Second World War, and the disclosure of detailed information on the products of German manufacturers, there was a pressing need for a revised up-to-date edition. In collaboration with the American Association of Textile Chemists and Colorists, the Society of Dyers and Colorists began this mammoth task that took over a decade to complete. It was recognized that the Second Edition of the Colour Index had three functions: (i) to correlate the different commercial names of individual dyes, (ii) to provide information on the mode of application, fastness properties, solubility, and so on, and the established usage of current commercial dyes, (iii) to provide information on the chemical structure and components, and mode of preparation of technical dyes, and to supply relevant literature and patent references.

The first volume of the new edition was published at the end of the Perkin Centennial year (1956). On completion, Part 1, the first two volumes, covered Technology and Commerce. Dyes and pigments were classified according to usage and were arranged in spectral hue order. Each colorant was given a usage name and number, known as the C.I. Generic Name (CIGN), e.g., C.I. Acid Yellow 23 and C.I. Pigment Blue 15. Methods of application, usage, and fastness properties were also included. Part 2, volume 3, covered the classification of dyes and pigments

according to their chemical constitution; an index of intermediate compounds was also given. Each colorant was allocated a five-figure number, known as the C.I. Constitution Number (CICN), e.g., C.I. 19,140 (for C.I. Acid Yellow 23) and C.I. 74,160 (for C.I. Pigment Blue 15). Methods of manufacture, inventors, and literature references were provided. Part 3, volume 4, consisted mainly of an alphabetical list of commercial names together with CIGNs and CICNs; cross-references to numbers used in the First Edition of the Colour Index and in the final edition of Farbstofftabellen were included. Overall, the new edition covered over 3500 unique colorants; a Supplement was published in 1963, an important feature of which was the introduction of C.I. Reactive Dyes as a new usage group. The new edition was updated by the publication of a series of Additions and Amendments, mostly quarterly, between late 1958 and the end of 1970.

Third Edition

Publication of the Second Edition established the Colour Index as the leading reference work for the classification of colorants and, by the end of 1963, copies had been supplied to over 50 countries. Collation of information for the publication of a Third Edition of the Colour Index began during 1965 and, again, there was collaboration between the SDC (Editorial Board) and the AATCC (Editorial Committee), a formal agreement being signed during 1969. The completely revised and enlarged Third Edition was published in five volumes at the end of 1971. The first three volumes (Part 1) cover technical information and classification by usage and volume 4 (Part 2) deals with chemical information and classification by structure; this volume also contains an alphabetic Intermediates Index and an Elementary Formula Index. At the time, volume 4 contained current as well as historical structural information. Commercial information is found in volume 5 (Part 3) under a Generic Names Index (which includes relevant C.I. Constitution Numbers) together with an alphabetic index. As with the Second Edition, the various volumes are paginated with four-figure numbers, the first digit in each number

indicating the volume in which that page is to be found [2].

The technical information provided in Part 1, listed under C.I. Generic Names, follows the alphabetical usage (application) sequence Acid Dyes, Azoic Colorants, Basic Dyes (volume 1), Developers, Direct Dyes, Disperse Dyes, Fluorescent Brighteners, Food Dyes, Ingrain Dyes, Leather Dyes (volume 2), Mordant Dyes, Natural Colorants, Oxidation Bases, Pigments, Reactive Dyes, Reducing Agents, Solvent Dyes, Sulfur Dyes, Condense Sulfur Dyes, and Vat Dyes (volume 3). Within each dye class, the hue order is yellow, orange, red, violet, blue, green, brown, and black; in the case of pigments, the list is extended to include white and metal. Where known, the relevant C.I. Constitution Numbers are given. The hue or shade is defined according to a Colour Index Hue Indication Chart.

The structural information provided in Part 2 (volume 4) is listed under C.I. Constitution Numbers (CICNs) contained within various chemical classes; related C.I. Generic Names (CIGNs) are included where appropriate. Each essential colorant is allocated a separate five-figure CICN. Where dyes or pigments differ only in the metal or acid used for salt formation, a subdivision has been made by addition of a sixth figure after a colon. The absence of a CIGN signifies a colorant of historical significance or a product no longer manufactured (some of these examples have been allocated CIGNs subsequently). In each category, the various colorants are arranged strictly on the basis of their chemical structures; consequently, dyes and pigments that are chemically related, but belong to different application classes, can be found together.

The main structural categories used to classify colorants in volume 4 (with CICN ranges shown in parentheses) are Nitroso (10000–10299), Nitro (10300–10999), Monoazo (11000–19999), Disazo (20000–29999), Trisazo (30000–34999), Polyazo (35000–36999), Azoic (37000–39999), Stilbene (40000–40799), Carotenoid (40800–40999), Diphenylmethane (41000–41999), Triaryl methane (42000–44999), Xanthene (45000–45999), Acridine (46000–46999), Quinoline (47000–47999),

Methine and Polymethine (48000–48999), Thiazole (49000–49399), Indamine (49400–49699), Indophenol (49700–49999), Azine (50000–50999), Oxazine (51000–51999), Thiazine (52000–52999), Sulfur (53000–54999), Lactone (55000–55999), Aminoketone (56000–56999), Hydroxyketone (57000–57999), Anthraquinone (58000–72999), Indigoid (73000–73999), Phthalocyanine (74000–74999), Natural Organic Coloring Matters (75000–75999), Oxidation Bases (76000–76999), and Inorganic Coloring Matters (77000–77999). The final category contains both natural and synthetic products, classified under their metals and listed alphabetically in the order metal (in a few cases), oxides, hydroxides, and then other compounds in anion sequence. In the case of compounds containing more than one metal, the listing is under the metal mainly responsible for the coloring properties of the product.

Incorporated within many of these ranges of organic colorants are further structural subclasses. The description of each class of colorants is preceded by a brief account of the characteristics of its members. Typical structures are usually shown and, where the class comprises a variety of chemical types, the order in which these are dealt with is explained. Within a particular class, simpler colorants precede more complex structures, with aliphatic derivatives appearing before aromatic compounds, which follow the order benzene, naphthalene, anthracene, and heterocyclic types. The colorants in each chemical class are usually arranged accordingly as they contain amino, hydroxyl, or both amino and hydroxyl groups. The arrangement within the various subclasses is, in general, in the substituent sequence halogen, nitro, acylamino, carboxamido, sulfonamido, and sulfonyl. In the case of azo colorants, compounds devoid of carboxylic acid and sulfonic acid groups are separated from those containing these substituents in order to bring together usage groups such as Disperse Dyes, Solvent Dyes, and Pigments which otherwise would be widely scattered. More comprehensive details of chemical progression are provided at the start of each class or subclass, as appropriate.

The first hard copy supplement to Volumes 1 to 4 was published in 1975, as volume 6, together with the first revision of volume 5. A supplement to Volumes 1 to 4 and 6 was published in 1982, as volume 7, together with the second revision of volume 5. A supplement to Volumes 1 to 4, 6, and 7 was published in 1987, as volume 8, together with the third revision of volume 5. Finally, a supplement to Volumes 1 to 4, 6, 7, and 8 was published in 1992, as volume 9, entitled *Colour Index International Third Edition (Fourth Revision)*, which also incorporated the latest revision of volume 5. Much of the new information had been published in a new series of quarterly *Additions and Amendments* updates from 1971 (Number 1) to, eventually, 2000 (Number 116).

Pigments and Solvent Dyes Edition

For the benefit of the non-textile pigment-using industries, a separate *Pigment and Solvent Dyes* extract, taken from Volumes 2, 3, 4, 5 (second revision), 6, and 7 of the Third Edition of the *Colour Index*, was published in 1982; this extract also included information on *Fluorescent Brighteners* having non-textile uses. A Supplement was published in 1989. An entirely new *Colour Index International Pigments and Solvent Dyes Edition* was published in 1997, followed by a Supplement in 1998. The new edition lists current *Pigments* and *Solvent Dyes* by C.I. Generic Name in hue sequence. Each new-format entry is in tabular form, headed, in most cases, by a structure and its relevant C.I. Constitution Number, together with indications of Hue, Chemical Class, First known product, CAS Number, EU Number, and CPMA Number, as appropriate. Commercial products are then listed, with an opportunity for manufacturers to indicate the physical form and the main uses (paint, plastics, printing inks, and other) for their product. In addition, a sentence can be added to indicate anything special, which may distinguish the manufacturer's product from others under that CIGN. The volume also has a comprehensive alphabetical list of commercial products, including company code letters, physical form, CIGNs, and CICNs.

Heritage Edition

Despite the usefulness of the *Pigments and Solvent Dyes Edition*, it was recognized that the future of the *Colour Index* lay in the eventual publication of a Fourth Edition online. In response to requests for historical information about colorants and their classification, the SDC published the *Colour Index Heritage Edition*, a DVD compilation of all the available information published in the various editions of the *Colour Index* between 1924 and 2000; the DVD contains over 17,500 pages of archive information presented in Portable Document Format (PDF) as scanned page images and was published in 2005.

Current Aspects

The Fourth Edition of the *Colour Index International Online* has continued with the various improvements incorporated into the 1997 *Pigments and Solvent Dyes Edition* [3].

Part 1 covers pigment and solvent dye classes and Part 2 covers dyes and related products. The new format has been devised through consultation with the coloration industry, including manufacturers, users, the Ecological and Toxicological Association of Dyes and Organic Pigments Manufacturers (ETAD), and the Color Pigments Manufacturers Association, Inc. (CPMA). Subscribers can search the online database under various categories, such as CIGN and CICN, chemical class, hue/shade, discoverer, first known product, CAS/EU number, manufacturer name, and commercial product name.

Entries are organized as Fingerprint pages, which aim to include information on C.I. Generic Name, C.I. Constitution Number, chemical structure, chemical class, shade, discoverer, first known product, CAS registry number, EU number, and classical name. Following the Fingerprint entry, commercial information is listed under the headings commercial name, manufacturer, physical form, main uses, and additional product details provided by the company.

Definitions

Revised and agreed definitions have been published on the SDC website as follows.

Essential Colorant(s)

Many commercial products that are placed on the market are not just 100% of the substance responsible for the color. They usually contain, in addition, quantities of other substances (generally referred to as additives) that improve the product's application properties: e.g., its dispersibility, flow, or flocculation resistance. Quite significant amounts of these additives may be present. The essential colorant is a dye or pigment responsible for the color of the product in the absence of additives.

ETAD Definition of Dyes

Dyes are intensely colored or fluorescent organic substances only, which impart color to a substrate by selective absorption of light. They are soluble and/or go through an application process which, at least temporarily, destroys any crystal structure by absorption, solution, mechanical retention, or by ionic or covalent chemical bonds.

CPMA Definition of Pigments

Pigments are colored, black, white, or fluorescent particulate organic or inorganic solids which usually are insoluble in, and essentially physically and chemically unaffected by, the vehicle or substrate in which they are incorporated. They alter appearance by selective absorption and/or by scattering of light. Pigments are usually dispersed in vehicles or substrates for application, as for instance in the manufacture of inks, paints, plastics, or other polymeric materials. Pigments retain a crystal or particulate structure throughout the coloration process.

Definition of C.I. Generic Names (CIGNs)

A C.I. Generic Name describes a commercial product by its recognized usage class, its hue, and a serial number (which reflects the chronological order in which related colorant types have been registered with the Colour Index). This definition enables a particular commercial product to be classified along with other products whose essential colorant is of the same chemical constitution and in which that essential colorant results from a single chemical reaction or series of reactions. The definition also includes multi-

constituent substances but excludes products obtained simply by physical admixture of essential colorants. In some cases, the same essential colorant is present in more than one usage category. Typically, a disperse dye can often be applied as a solvent dye, and a vat dye can sometimes be used as a pigment, e.g., C.I. Pigment Blue 60 and C.I. Vat Blue 4.

Colorants listed under a C.I. Generic Name whose serial number contains a colon differ slightly from those listed under the parent CIGN. This slight difference is usually chemical but may be due to different crystal modifications in the case of certain pigments. Historically, a colon number was added to the serial number of a particular C.I. Generic Name when a new product was reported by a manufacturer as being similar to the original. Similar to can generally be interpreted as slightly different chemically. This imprecise definition gave manufacturers the opportunity to request new C.I. Generic Names for products that should have been given colon numbers. Conversely, subsequent disclosures have sometimes revealed that certain products issued with colon numbers should really have been allocated new CIGNs.

Definition of C.I. Constitution Numbers (CICNs)

In the Third Edition of the Colour Index, each essential colorant disclosed for publication was classified and allocated a separate five-figure C.I. Constitution Number. The original numbering of colorants was in multiples of five so as to leave space for future structurally similar dyes and pigments to be assigned numbers in the correct chemical sequence. From 1997, however, new disclosures have been allocated six-figure C.I. Constitution Numbers (before any additional colon number, where relevant) in order to alleviate the congestion experienced in certain areas, notably monoazo colorants.

Organic Structures

The chemical constitutions depicted in volume 4 (1971) are primitive, uninformative, and

inaccurate; no account is taken of aromaticity, tautomerism, or hydrogen bonding. More realistic structural representations did not appear until the publication of volume 9 (1992). All the chemical structures disclosed in the *Pigments and Solvent Dyes Edition* (1997) were revised and updated [4]. In particular, dominant ketohydrazone tautomers are given for derivatives of phenylazonaphthols and for acetoacetanilide and pyrazolone colorants. Wherever possible, structures are drawn in conformations dictated by the requirements of hydrogen bonding. The positive charge associated with the triarylmethane colorants is located at the central carbon atom for convenience. This publication marked the start of a complete revision of all the structures currently appearing in the online database.

Benzene rings are shown as Kekulé structures. Although the circle in a hexagon notation is a useful shorthand representation of the aromatic sextet in benzene, it is simplistic. Apart from benzene itself, bond type alternation occurs in all its derivatives except for fully symmetrical structures such as hexafluorobenzene. Fusion between a benzene ring and a heterocyclic ring system is always shown at a double bond. For naphthalene, the central bond is always shown as a double bond, in accordance with reality – alternating bond type occurs in naphthalene itself and in all its derivatives. With ketohydrazone structures, the uninvolved ring retains its three double bonds.

The central ring of anthraquinone and its derivatives possesses quinonoid character, so that the Kekulé-type terminal benzene rings are shown as mirror images. As far as possible, a similar approach is used with larger polycyclic quinones.

Azo bonds are never linear. The intermediates used to form azo metal-complex dyes are depicted as hydroxyazo structures.

Use of Colon Numbers

In the Third Edition (1971), the so-called colon numbers were used to subdivide both CIGNs and CIGNs in order to distinguish minor differences in properties or structure, typically between the

variable salts of anionic and cationic dyes or pigments. For example, C.I. Acid Orange 7 (C.I. 15510) is the sodium salt and C.I. Pigment Orange 17 (C.I. 15510:1) is the barium salt. Similarly, C.I. Basic Violet 1 (C.I. 42535) is the chloride salt and C.I. Pigment Violet 3 (C.I. 42535:2) is the phosphotungstomolybdic acid (PTMA) salt.

The use of colon numbers has been extended, somewhat arbitrarily, to include small structural differences, such as ethyl instead of methyl; the adoption of six-figure CIGNs has made it possible to minimize such usage. Improved solubility in nonaqueous solvents can be achieved by forming amine salts of azo metal-complex dyes and, in such cases, a colon number has been added to the CIGN of the parent dye. In a similar manner, solubility in nonaqueous solvents can be achieved with some phthalocyanine derivatives where the same chromophore is involved but the application properties differ considerably. For example, treatment of C.I. Direct Blue 86 (C.I. 74180) with 2-ethyl- *N*-(2-ethylhexyl)-1-hexanamine gives C.I. Solvent Blue 129 (C.I. 74180:2); the barium salt (C.I. 74180:1) is C.I. Pigment Blue 17.

Colon numbers have also been used to distinguish between various crystal modifications and analogues of metal phthalocyanines. For example, C.I. Pigment Blue 15:1 (C.I. 74160), which is reddish-blue, is the α -modification and C.I. Pigment Blue 15:3 (also C.I. 74160), which is greenish-blue, is the β -modification. The only structural difference between C.I. Pigment Blue 15 (C.I. 74160) and C.I. Pigment Blue 75 (C.I. 74160:2) is a change in the central metal atom from copper to cobalt.

Relatively few colon numbers have been issued in the inorganic pigments section of the *Colour Index*. For some extended pigments obtained by a process of coprecipitation, or physical admixture, both the CIGN and the corresponding CIGN have been given a colon number. For example, the cadmium sulfide pigment C.I. Pigment Yellow 37 (C.I. 77199) is coprecipitated or admixed with barium sulfate to form C.I. Pigment Yellow 37:1 (C.I. 77199:1). Other limited usage has been varied and

somewhat inconsistent, and there is no intention of introducing any new colon number usage into the inorganic pigments section.

Product Registration

The allocation of a new C.I. Generic Name for inclusion in the online database is made in response to a specific request from the manufacturer of the new commercial product. As well as providing standard usage information, the manufacturer also has the opportunity to highlight any special features of the product.

In order to minimize the possibility of duplication, the chemical structure of the essential colorant has to be disclosed but will be kept confidential if so desired; for each new product, an appropriate C.I. Constitution Number is allocated.

Cross-References

- [Chromophore](#)
- [Colorant, Textile](#)
- [Coloration, Fastness](#)
- [Coloration, Mordants](#)
- [Coloration, Textile](#)
- [Dye](#)
- [Dye, Functional](#)
- [Dye, Metal Complex](#)
- [Pigments, Organic](#)

References

1. Burdett, B.C.: The colour index: the past, present and future of colorant classification. *J. Soc. Dye. Colour.* **98**, 114–120 (1982)
2. Shore, J.: Classification and general properties of colorants. In: Shore, J. (ed.) *Colorants and auxiliaries organic chemistry and application properties colorants*, vol. 1, 2nd edn, pp. 1–44. Society of Dyers and Colourists, Bradford (2002)
3. Hallas, G., Renfrew, A.H.M.: Representation of tautomeric colorants in the colour index. *J. Soc. Dye. Colour.* **112**, 207–208 (1996)
4. <https://colour-index.com>

Colouring

- [Coloration, Textile](#)

Combustion Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
Netherlands

Synonyms

[Flame light lamps](#)

Definition

Lamps that produce light as a result of an exothermic reaction between the vapor of a solid, liquid, or gaseous fuel, consisting of hydrocarbons and oxygen.

Types of Combustion Lamps

Torches, oil lamps, candles, and gas lamps all are combustion types of lamps. The light comes from the flame that is the result of a reaction between oxygen and the vapor of a solid fuel in the case of candles, of a liquid fuel in the case of oil lamps, and of a gaseous fuel in the case of gas lamps. A spark is needed for starting the reaction.

Working Principle

In all combustion types of lamps, after a spark has initiated the process, the combustion reaction takes place between the gaseous state of hydrocarbons of the fuel and oxygen from the air. The products of the reaction are carbon dioxide (CO₂), water vapor



Combustion Lamp, Fig. 1 Flame of a candle

(H₂O), heat, and light in the form of a flame. In oil lamps and candles, the wick draws, by its capillary action, the fuel to the flame against the force of gravity. It is not the liquid part of that fuel that takes part in the combustion reaction but the evaporated fuel around the wick. With all combustion lamps, carbon particles of high temperature (soot), resulting from incomplete combustion, are brought to incandescence and contribute to the total light radiation. The blue light at the bottom of the wick (Fig. 1) is the result of the combustion reaction; the yellowish part of the flame is the result of incandescent light from the carbon particles.

The combustion reaction vitiates the atmosphere by consuming oxygen and returning carbon dioxide (CO₂). Non-complete combustion leads to the emission of harmful gasses like carbon monoxide (CO), sulfur oxides (SO_x), and nitrogen oxides (NO_x).

Oil Lamps

Oil lamps represent the oldest form of artificial light. Scraped-out stone oil reservoirs (Fig. 2) with a wick have been in use for lighting purposes since prehistoric times some 20,000 years ago [5, 6]. The basic conception of oil lamps has



Combustion Lamp, Fig. 2 Stone oil lamp found in Lascaux, France, 17000 BC. (Photograph: Sémhur, Creative Commons 3.0 unported)

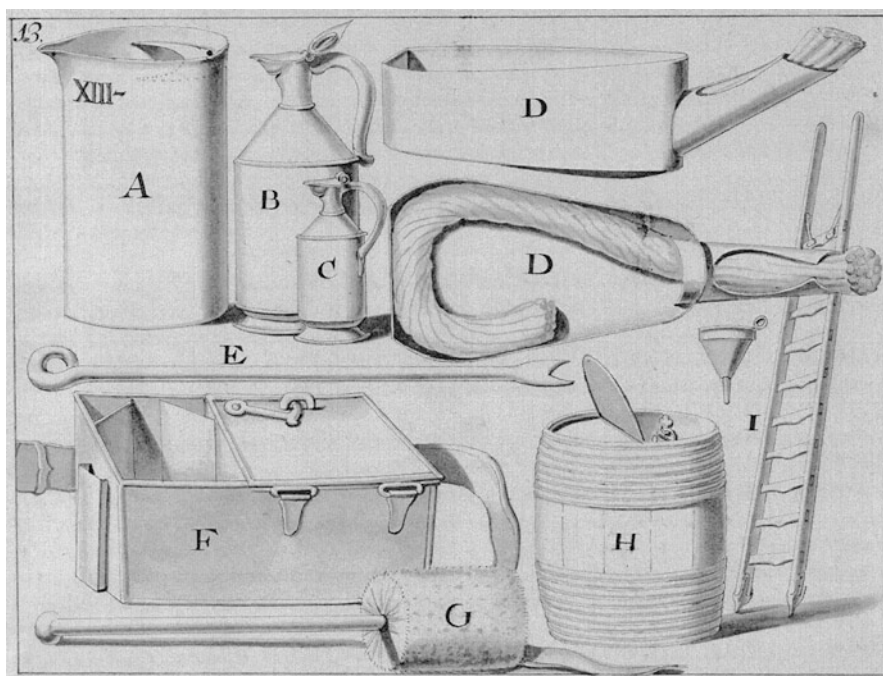
remained unchanged although the materials used and the type of construction have been advanced quite a bit. The Dutchman Jan van der Heyden, for example, developed in 1663 a closed oil reservoir for street lanterns [7]. He not only made a detailed description of the production process of the oil lamp and lantern but also of the set of maintenance material the lamplighter crew needed (Fig. 3).

The industrial revolution some 200–250 years ago asked for artificial lighting in industrial premises, and in that period a boom in new developments in the technology of oil lamps is seen, followed by developments of entirely new lighting products, such as gas lighting and later electrical lighting. Until the end of the nineteenth century, oil lamps have been in general use, especially for domestic lighting.

Materials and Construction

The Oil

Vegetal or animal oil, rich in carbon, is used. The type was dependent on the availability in the actual region. In warm areas vegetal oil was made from olives, coconuts, and palms and in more moderate climate regions, from colza, linseeds, and rapeseeds. Animal oil was obtained from fish, whale, or domestic animals. Especially sperm whales were hunted and slaughtered for oil obtained from their head cavities. Around 1850 kerosene (also referred to as paraffin oil) was produced from crude oil (also referred to as



Combustion Lamp, Fig. 3 Maintenance set for a lamplighter of oil street lighting installations around the end of the seventeenth century (Drawing van der Heijden 1663, [7])

petroleum) through a refining distillation process. Quickly it became the standard fuel for oil lamps which became known under the names of paraffin, kerosene, or petroleum lamps.

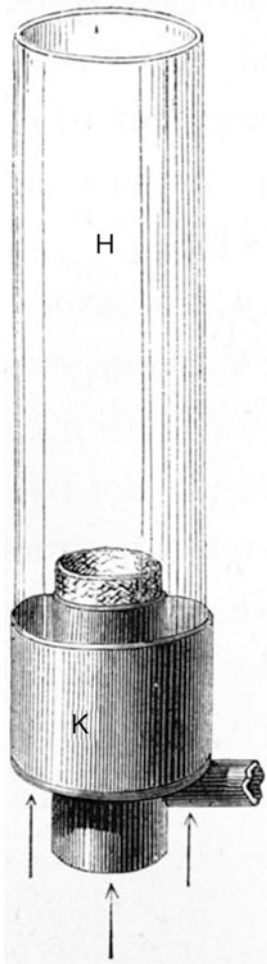
The Wick

In the early oil lamp days, the wick was made from bark, moss, or plant fibers that were twisted together. Later cotton was usually used for wicks. Some oil lamps had multiple wicks, up to 16 in certain Greek and Roman types. Around 1780, experiments with the shape of the wick resulted in a burner that consisted of two concentric tubes between which a tubular wick is located (Fig. 4). Through the open central tube, air is drawn so that the combustion is improved and consequently the light output as well (equally to some ten candles) while reducing smoke and smell. This oil lamp type is named, after its Swiss inventor, the Argand lamp. A further improvement of this type constituted of a glass chimney placed over the flame that increased the upward air draft through the hollow tube (see Fig. 4 again).

The Fuel Reservoir or Lantern

Originally the fuel reservoir was a simple open tray made out of stone, seashell, or earthenware in which the wick was free-floating or laid in a groove in the rim of the tray. Later the reservoir was provided with a nozzle or spout through which the wick was led. Sometimes underneath the spout, a gutter was mounted to catch spilt oil to avoid pollution and for reuse (Fig. 5).

The material used for the reservoir becomes gradually more advanced: brass, copper, silver, pewter, glass, or porcelain. The reservoirs made out of these materials are closed usually with a lid, although with some versions, filling of the reservoir had to be done via the spout. Around 1,800 double-reservoir oil lamps came into use on the initiative of, again, Argand who earlier introduced the double concentric tube burner. These lamps had the advantage that the supply of oil to the wick was relatively constant from a small reservoir that continuously was filled by the force of gravity by a secondary larger reservoir fitted somewhat higher (Fig. 6).



Combustion Lamp, Fig. 4 Argand oil burner with a tubular wick inside a hollow cylinder equipped with a glass chimney [2]

The Carcel lamp had a clockwork that operated a pump to raise the oil to the wick. The light output of the Carcel lamp was so constant that its horizontal intensity was for some time used as the unit of intensity. That one Carcel equals approximately 9.8 cd illustrates the high light output of the Carcel lamp relative to that of a candle. The less complicated “moderator lamp” with a spring-loaded piston to pressure-feed the burner became popular for general lighting purposes. The last improvement in oil lamps dates from around 1900 and in fact comes from technology then already in use for gas lamps. It combines the use of a gas mantle (see a further section



Combustion Lamp, Fig. 5 Brass spout oil lamp with a gutter to catch spilt oil



Combustion Lamp, Fig. 6 Double-reservoir oil lamp



Combustion Lamp, Fig. 7 A modern, double-mantle, pressurized kerosene lantern

“Gas Lamps”) with a hand pump that pressurizes the fuel liquid to force it into the burner for better combustion where its flame brings the mantle to incandescence. The resulting light is brighter and has a cooler color. These types of lamps are still produced today for emergency lighting purposes and for outdoor use (Fig. 7).

Properties

Simple oil lamps have a lumen output of about 10 lumen and a luminous efficacy (based on heat release) of some 0.1 lm/W [1, 3]. The mixture of light from incomplete combustion and incandescence of carbon particles results in a correlated color temperature of approximately 2,000 K [8].

Argand and moderator type of oil lamps have a lumen output in the range of 50–200 lumen

(comparable to 5–25 W incandescent lamp) with an efficacy of 0.1–0.3 lm/W.

Oil lamps equipped with a gas mantle may raise the lumen output to more than 500 lumen with an efficacy of 0.5–1 lm/W (more than 1.5 lm/W for the pressurized types). The correlated color temperature increases to some 2,700 K.

Candles

Candles came in use much later than oil lamps. Spillage of oil and the associated risk of fire have always been a problem with oil lamps. With the invention of the candle made of non-fluid material, the spillage problem was solved albeit not the risk for fire. The candle was less fragile than the oil lamp and therefore more easily portable. The Romans, from the time of the birth of Christ onward, have been responsible for the dissemination of the candle throughout Europe and the Middle East [6]. Today candles are mainly used for devotion and for ambience lighting. Annual retail sales in the USA of candles, today, exceed five billion pieces (calculated based on [1]).

Materials and Construction

The Candle Substance

The first candles were made of hard animal fat (tallow) or of beeswax. Wax candles were of much better quality but also much more expensive. From the end of the eighteenth century, the use of relatively cheap fat from the spermaceti organ in the head of sperm whales improved the quality of candles. Standardized candles on this basis were used as standards for light intensity: the candle-power (only in 1948 the SI unit candela came into use with a value roughly equal to one candle-power). Around 1830 stearin, obtained by chemical treatment of animal or vegetal fat or oil, was added which increased the melting point and consequently reduced dripping of candle fat. Around 1850 candles became much cheaper because of the

use of solid paraffin, a distillation product from crude oil. At that time, however, gas lighting was already, at many places, the source of lighting.

Candles are produced in some different ways:

- By dipping the wick repeatedly in molten candle substance
- By pouring molten candle substance in a glass container
- By pouring molten candle substance in molds
- By drawing soft candle substance through a hole (machinal)

The Wick

The first candles had a wick made of a piece of wood, cord, or animal skin. Around 1800 the braided cotton wick was introduced that reduced the disturbing smoke that accompanied burning candles. The braided wick can have a stiff core, originally made of lead and later of zinc, of paper, or today of synthetic fibers. Most wicks are impregnated with wax to facilitate ignition. Early wicks had to be trimmed regularly. Later wicks got such a structure that they bend, and their residues dip into the molten fuel and are completely consumed so that trimming is not needed.

The Lantern

To reduce the risk of fire but primarily to enable the use of candles outside, candles were used in lanterns made of a perforated metal sheet or with windows of animal horn or glass. Only from the eighteenth century onward, candle lanterns sometimes were equipped with mirrors or lenses to concentrate the light in specific directions.

Monumental buildings and homes of the rich were lit with luxurious candle chandeliers, sometimes decorated with pieces of cut glass that made the light sparkle. In contrast to oil lamps, candles were only for a very limited period used for street lighting.

Properties

The light and color properties of the candle are similar to those of simple oil lamps. The lumen

output of a candle flame is approximately 10 lm and its luminous efficacy 0.1 lm/W [1, 3]. The correlated color temperature is around 1,900 K [8].

Gas Lamps

Oil lamps and candles are light sources where the energy or fuel is stored in the lamp itself. Gas lamps were the first lamps where the energy is distributed to the lamps from a central energy depot at a centralized remote location. The challenge was not only to develop the lamp itself but also the development of town-sized gas production and distribution systems. In 1785, Dutch Jean-Pierre Minckelers demonstrated the use of coal gas to produce light by lighting his lecture room at the University of Leuven with gaslight. The first to make a public demonstration of gas lighting was the Scot William Murdoch when he in 1802 installed a gas pipe network with gas burners for the lighting of the facades of a range of buildings of James Watt's Soho Factory in Birmingham [3, 5, 6]. Oil and candle lighting was quickly replaced by gas lighting, first in street lighting and industrial premises, quickly followed in indoor areas of the rich. In 1875 the new Paris Opera got some 45 km of gas pipe with 960 gaslights connected to it.

With the introduction of electric incandescent lamps in 1880, some 75 years after the first use of gas lighting, the use of gas lighting decreased quickly. Even today, however, in some cities in the Western world, gas lighting for street lighting is still in use. The city of Berlin, for example, uses some 40,000 gas lanterns.

Materials and Construction

The Gas

Gas for lighting was sometimes produced by heating animal fats or vegetal oil but mostly by heating coal. When the latter is done in an air-free atmosphere, in cast-iron retorts, the percentage of methane in the gas is higher, resulting in better combustion and therefore in more light. After its



Combustion Lamp, Fig. 8 Bicycle carbide lantern working on acetylene gas

original application, this type of gas was named “illuminating gas.” Illuminating gas produces a whiter light than the blue light of natural gas (only after the invention of the gas mantle – see next section – also natural gas could be used for gaslights). The gas was stored in huge gas storage tanks or gasometers, in some cities preserved as historic landmarks.

A special kind of gas was used for carbide gas lamps on bicycles and early automobiles: acetylene gas obtained by dripping water from a small reservoir in another reservoir under it that contains calcium carbide (Fig. 8).

The Burner

Open Flame Burner The first burners, so-called open flame burners, were merely a hole or a series of holes at the end of a pipe. The double concentric tube burner with a glass chimney, as used since the end of the eighteenth century by Argand for his oil burners, was often used for open flame gas burners as well. It was realized that the higher the temperature of the flame, the larger the lumen output of the flame would be. In 1858 William Sugg, therefore, introduced burners made from

non-heat-conducting steatite that became hotter than heat-conducting metal burners.

Gas Mantle Already in 1825, it was known that by putting solid material into the flame, this material could be brought to incandescence. In this way, both the light output and color of the flame could be influenced. A piece of lime from limestone brought into the flame resulted in an extremely intense and compact light source that was used as a spotlight in theaters (Fig. 9). The expression “in the limelight” originates from this.

In 1885 the Austrian Carl Auer von Welsbach succeeded in using a knitted mantle of solid material around the gas flame (Fig. 10). The material consists of a pear-shaped net of fabric impregnated with the nitrates of the rare-earth metals thorium and cerium. After drying, the fabric is burned off and what remains is a fragile, inflammable structure: the gas mantle existing of the oxides of thorium and cerium that is brought to incandescence in the gas burner. Because of its fragility, the mantle has to be replaced regularly.

The gas mantle improved both the light output and efficacy of gas lamps considerably because thorium and cerium oxides produce more light in the visible spectrum range than a black body at the same temperature does (selective radiation). The gas mantle invention has stretched the actual use of gas lighting well into the era of the first 50 years of electrical lighting.

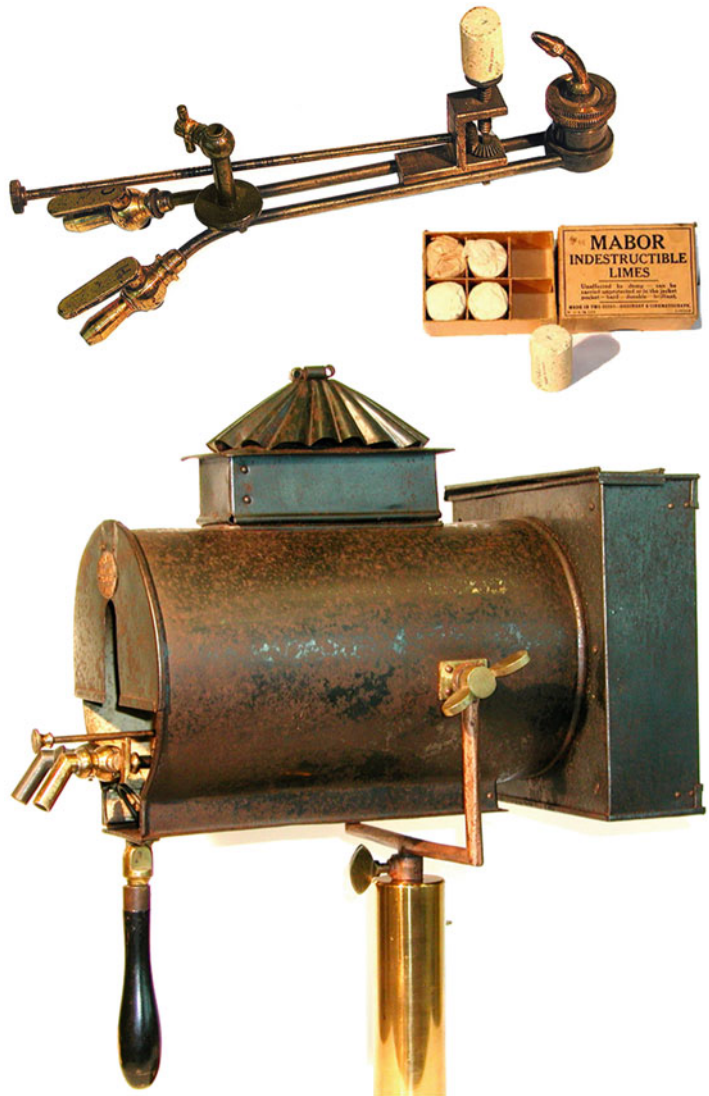
Regenerative Burners Around 1880 the so-called regenerative gaslight systems further improved the efficiency of gas lamp systems. The heated air produced by the flame is guided such that it preheats the incoming air needed for combustion. Often this principle was applied with inverted gas mantle burners in which the heat of the flame is better retained in the mantle.

The Lantern

The design of the early gas lanterns was very much based on the design of oil lanterns of that period.

Combustion Lamp,

Fig. 9 From top to bottom: limelight gas burner with a piece of limestone, a box with limestones, theater spotlight projector in which the limelight gas burner is fitted. (Photographs: Stage Lighting Museum, Israel, creator Dan Redler)



Chandeliers for gas lighting were equipped with pull chains for switching the gas supply on and off and to control, through the quantity of gas supply, the light output. They also often had thin secondary gas pipes connected to the so-called pilot lights that were always on so as to enable the main gaslights to be ignited without the use of matches (Fig. 11).

Lanterns for gaslights with gas mantle always had a glass cover to protect the fragile mantle and to restrict glare. The glass cover, of course, was open at the bottom to permit inlet of air needed for

combustion. The top part of the lantern had a ventilation shaft working as a chimney (Fig. 12). Most lanterns for use in street lighting and for the lighting of industrial premises had multiple gas mantle burners varying in number from 2 to more than 10.

Ignition Control

In large part of the nineteenth century, gas street lighting lanterns were ignited each day at dusk by a patrolling lamplighter (Fig. 13). With a stick on

which a torch was fitted, he opened a trap door in the bottom of the lantern, opened the gas valve with a hook on the stick, and ignited the flame with the torch.

Later, gas lanterns had for remote ignition a pilot lamp connected through a bypass to the main gas pipe. Gas supply was remotely controlled by

gas valves actuated by gas pressure or by a mechanical clockwork built into the lantern. From 1900 onward remote gas igniters came on the market which did not need a pilot light. A platinum sponge, existing of a porous mass of finely divided platinum, initiates a flame by catalytically combining oxygen with hydrogen from the gas when the gas valve is opened. Later, also battery-operated remote igniters came into use.

Properties

A single gas burner without mantle has a lumen output of 10–50 lumen with an efficacy of 0.2–0.5 lm/W [1, 3]. The spectrum of radiation depends on the type of gas and the oxygen supply that together determine the quality of the combustion reaction and the soot particles that take part in the incandescence process. Figure 14 shows different colored gas flames resulting from different types of gas and different oxygen supply. Open flame burners on “illuminating” gas have the color of flame as shown on the left: yellowish white light from incomplete combustion and incandescence of soot particles. Modern gas ovens have



Combustion Lamp, Fig. 10 Gas mantle

Combustion Lamp, Fig. 11 Gas lamp chandelier with gas mantles and with thin secondary gas pipes connected to pilot lights for ease of ignition. (Photograph: Pat Cryer, taken at the Museum of Welsh life, Cardiff, Wales)



Combustion Lamp,

Fig. 12 Multiple gas mantle street lighting lanterns as in use in 2012 in Dusseldorf, Germany. (Photograph: Wout van Bommel)



flames as illustrated on the right: a blue flame from complete combustion, without soot particles taking part in the process. The correlated color temperature of the open burner gas flame, burning on illuminating gas, lays around 2,400 K [4].

Gas burners with a gas mantle have per mantle a lumen output of 200–600 lumen (comparable to that of a 25–60 W incandescent lamp) at 1–2 lm/W [1, 3, 8]. It is interesting to compare this luminous efficacy with that of Edison's first incandescent bulb: 1.4 lm/W. The correlated color temperature is dependent on the composition of the mantle.

Typical values are 2,700–2,900 K [4]. The lifetime of a gas mantle was some 50–200 h. Modern gas mantles, as used today in professional street lighting lanterns, have a life of up to 4,000 h.

The effect of the main pressure on the performance of gas lamps was much less than the effect of main voltage on the performance of most electric lamps. From the last quarter of the nineteenth century, most gas lanterns are equipped with a small pressure governor that keeps the outlet pressure within acceptable limits.



Combustion Lamp, Fig. 13 Lamplighter. (Photograph: Klearchos Kapoutsis, Foter)



Combustion Lamp, Fig. 14 Different colors of gas flame depending on the type of gas and oxygen supply. From left to right: more complete combustion. (Photograph adapted from WarX: Creative Commons 3.0 unported)

Cross-References

► [Incandescence](#)

References

1. Dillon, S.E.: Characterization of candle flames. *J. Fire. Prot. Eng.* **15**, 265–285 (2005)
2. Figuiet, L.: *Les Merveilles De La Science, ou Description Populaire Des Inventions Modernes*. Hachette Livre, Paris (1891)
3. Luckiesh, M.: *Artificial Light Its Influence Upon Civilization*. The Century Books of Useful Science, New York (1920)
4. Macbeth, N.: Color temperature classification of natural and artificial illuminants. *Trans. Illum. Eng. Soc.* **23**, 302–324 (1928)
5. Rebske, E.: *Lampen Laternen Leuchten*. Franckh'sche Verlagshandlung, Stuttgart (1962)
6. Stoer, G.W.: *History of Light and Lighting*. Philips Lighting Division, Eindhoven (1986)
7. Van der Heijden, J.: *Het licht der lamplantarens*. Town Library, Haarlem (1663)
8. Waldram, J.M.: *Street Lighting*. Edward Arnold, London (1952)

Complementary Colors

Paul Green-Armytage
School of Media, Creative Arts and Social
Enquiry, Curtin University, Perth, Australia

Synonyms

[Contrary colors](#); [Contrasting colors](#)

Definition

Complementary colors are colors that “complete” each other. This completion can be understood in terms of some physical relationship or in terms of how the colors are related in their appearance. There are different ways of establishing these relationships, two being widely accepted: By one definition, two paints, inks, or colored lights are complementary if their mixture can yield a neutral black, gray, or white. This is a physical

relationship that can be demonstrated. By another definition, two colors are complementary if the afterimage of one color has the same hue as the other color. This is a phenomenal relationship that can also be demonstrated. It is also common simply to claim that colors opposite to each other on a color circle are complementary, without further explanation or justification. Complementary relationships can be helpful when mixing paints to produce particular results. Complementary relationships also feature in theories of color harmony.

Introduction

The notion of complementary colors is caught between science and art, between what can be measured and what cannot. But this notion can also be seen as a bridge. If there are objective ways of establishing complementary color pairs, and if such pairs are found to be pleasing, then complementary colors may be a key to color harmony – harmony could be subject to measurement. One difficulty here is that different ways of establishing complementary relationships do not yield exactly the same results. A number of these ways are described and illustrated in the sections that follow. There is also the problem of what comes first, the objective methods or the experience of harmony. The idea that some color pairs are more pleasing than others is older than any demonstration of particular physical or phenomenal relationships of the kind associated with complementary colors. Perhaps the earlier judgments are endorsed by the later demonstrations, or perhaps the demonstrations show color relationships that are now widely agreed to be harmonious.

Color Harmony

Harmony is a slippery word. Most definitions deal with harmony in music, but there it has more than one meaning. It can mean a combination of notes which have a pleasing effect [1]. It can also mean a combination of notes organized according to a system of structural principles [2]. The same definitions could be applied to color combinations. According to the first definition, only those color combinations

that are found to be pleasing are harmonious. This could make color harmony a private matter; harmony, like beauty, would be in the ears of the listener or the eyes of the beholder. Only some kind of consensus could establish a wider claim for some color combinations to be accepted as more harmonious than others. Whether the judgment is made by an individual or a group, this definition depends on evaluation. With the second definition, there are also limitations. Only with certain relationships between notes and colors can a combination be called harmonious. The relationships can be measured, so the definition depends on description. Philosophers might argue about the possibility or impossibility of any link between description and evaluation, but for most people, to say that there is an interval of a third between the notes C and E when they are sounded together in a chord is description, while to say that the chord is pleasing is evaluation. Similarly, to give Munsell or Natural Color System (NCS) notations for the two colors in a particular combination is description, while to say that the combination is beautiful is evaluation. The idea of a link between description and evaluation as the basis for some theories of color harmony is discussed in a paper that was presented at a conference in Gothenburg in 1996 [3].

Pleasing Color Combinations

Before the invention of the color circle, or any notion of a special relationship between colors that are opposite to each other on a circle, there was a recognition that certain color pairings are more satisfying than others. Leon Battista Alberti, in fifteenth-century Italy, claimed that “there is a kind of sympathy among colours whereby their grace and beauty is increased when they are placed side by side. If red stands between blue and green, it somehow enhances their beauty as well as its own” [4, p. 85]. Leonardo da Vinci introduced the notion of “contrary colors” and suggested that pairs of contrary colors enhance one another. “The colours which go well together are green with red or purple or mauve, and yellow with blue” [5, p. 73]. Martin Kemp points out that “this account of colours which ‘go well together’

comes close to the doctrine of ‘complementary colours’ as defined in the eighteenth and nineteenth centuries, but lacks the systematic base provided by the later colour wheels” [6, p. 284].

To find out whether people might intuitively choose particular colors as going well together, without appeal to a color wheel or any theories of complementary colors, a preliminary investigation was carried out at Curtin University of Technology in 1994. Students worked with colors of equal nuance from the Natural Color System (NCS). Each was assigned a particular color and asked to find a color to go with it so that the combination would be the most beautiful, most exciting, or most harmonious. The intention was to go back to the kind of judgments made by Alberti and Leonardo, to find colors that were most “sympathetic” or “contrary.” Students generally agreed on which color pairings were the most “beautiful” and the most “exciting”, but there were two schools of thought about what kind of color relationship is “harmonious” – colors of similar or contrasting hue. And it was not always the same colors that were selected as contrasting. There was a small range comparable to the “red or purple, or mauve” as nominated by Leonardo for “going well” with green. A more rigorous study, with a larger number of participants, might lead to firmer conclusions. But it does not seem likely that an approach, from this direction, to the identification of complementary colors would yield precise results. The “complementary” of a given hue would not be a single hue but a narrow range of similar hues. This can be set beside the imprecise results when the approach is from the other direction, starting with the theories.

Ways of Establishing Complementary Colors

With different ways of establishing complementary relationships yielding different pairs, no color can be said to have a single complementary hue unless a decision is first made about which way of establishing complementary relationships is “correct.”

Colored Shadows

In one of his diagrams, Leonardo da Vinci illustrates a spherical body (a) being lit by two lights (d) and (e). Two shadows are cast (b) and (c). He explains that “the shadow formed by the light *e*, which is yellow, will tend towards blue, because the shadow of the body *a* is formed on the floor at *b*, where it is exposed to the blue light, and accordingly the shadow made by the light *d*, which is blue, will be yellow at the location *c*, because it is exposed to the yellow light” [5, p. 74].

Later observers noticed that a blue shadow would be cast by a yellow light even if another light, shining on the shadowed area, was not blue but white. A systematic study of colored shadows was carried out by Count Rumford who reported his results to the Royal Society of London in 1794. Rumford concluded that the phenomenon was subjective. He went further and suggested that his results might serve as a guide to artists in “the magic of coloring,” and he introduced what was then a new term: “one shadow may with propriety be said to be the *complement* of the other” [7, p. 295]. Moses Harris, in 1766, also describes the colored shadow phenomenon and points out how the colors of the light and the shadow are opposite to each other on his color circle [8, p. 8]. Johann von Goethe explains how to conduct the colored shadow experiment with a candle and the full moon as light sources. And he has a beautiful account of how he experienced a sequence of colored shadows on a walk in the Harz Mountains. The ground and trees were covered in snow and hoar frost. Toward the end of the day, the shadows cast by the trees were “decidedly blue, as the illumined parts exhibited a yellow deepening to orange.” Then, as the sun began to set, it “began to diffuse a beautiful red colour over the whole scene around me, the shadow colour changed to a green, in lightness to be compared to a sea-green, in beauty to the green of the emerald. The appearance became more and more vivid: one might have imagined oneself in a fairy world, for every object had clothed itself in the two vivid and so beautifully harmonising colours” [9, pp. 34–35]. Colored shadows may be a subjective phenomenon, but it is possible to record them in photographs, as shown in Fig. 1.



Complementary Colors, Fig. 1 Demonstration of complementary colors as seen in colored shadows. At top left the candlestick is lit by two white lights and casts gray shadows. In the other photographs, colored filters were

placed in front of the light shining from the right while neutral filters were placed in front of the other light to equalize the illumination. The shadow colors can be described as being complementary

Afterimages

Colored shadows may be linked to the experience of afterimages. Martin Kemp identifies the Comte de Buffon as “the pioneer in the study of subjective colour.” Buffon described colored shadows and also afterimages at the French Academy of Sciences in 1742–1743 [7, pp. 294–295]. Afterimages were introduced by Moses Harris when he explains how the world will look if green-tinted spectacles are worn for about 5 min and then removed: “every scene and object will look of a fiery red, opposite to green you will find red” [8, p. 8]. In the Harris color circle, colored shadow and afterimage color pairs are to be found opposite to each other. For Goethe, afterimages were the key to color harmony. He saw, in the experience of an afterimage, the eye’s search for completeness, and he concludes that “in this resides

the fundamental law of all harmony of colours” [9, p. 317]. Goethe had only six colors in his color circle so, within each segment, there could be a range of different hues under the same name. Red would range from orange red to purple red and in that range the afterimage for a particular green might be found. If more precise identification of afterimage hues is required, there are two questions to be considered: If a particular red is the afterimage of a particular green, does it follow that the same green will be the afterimage of that red? Is the afterimage phenomenon reciprocal in this way? The second question takes account of afterimages as subjective experiences peculiar to the individual. So is it possible to know whether everyone experiences the same hue as the afterimage of a given color?

Afterimage Color Progression

The assumption that two colors could be each other's afterimage was investigated by Nathan Cabot Hale. Cabot Hale claimed to have found what he called "after-image color progression" [10, p. 262]. Having experienced a yellow as the afterimage of a blue, he found that the afterimage of the yellow was a violet. Then the afterimage of the violet was a green and this progression continued right round the circle. Cabot Hale's claim was investigated, in turn, by Western Australian artist Sally Douglas. Douglas did not find such strong hue shifts as those reported by Cabot Hale, but she did find progression all round the circle. Her method and results are shown in Fig. 2.

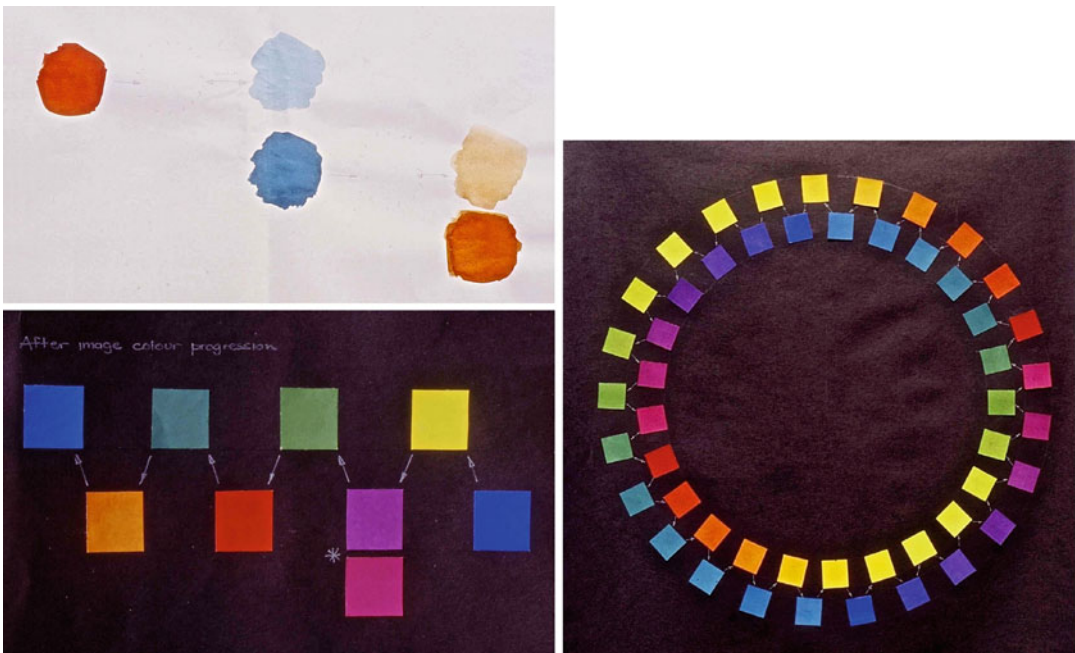
Having painted a patch of reddish orange, Douglas stared at it, looked away, and matched its afterimage which she saw as a pale turquoise. She then increased the intensity of the turquoise, being careful to retain its hue, and repeated the process for the turquoise. The afterimage of the turquoise was slightly more yellowish than the

original reddish orange. In a carefully controlled operation, where such color relationships were established in random order, she found, when she put it all together, a progression right round the circle.

Individual Variations in the Experience of Afterimages

To see if individuals have the same afterimage experiences from given colors, students at Curtin University of Technology were assigned specific paints and asked to match the colors they experienced as the afterimages. Some of the results are shown in Fig. 3.

The slight variations in hue may be due to inaccurate mixing but studies by Marian-Ortolf Bagley have confirmed that there are, indeed, variations in how individuals experience afterimages [11]. Rather than have the participants in her study match the afterimages in paint, she asked them to identify their afterimage colors from the samples in the Munsell book of color.



Complementary Colors, Fig. 2 Afterimage color progression as recorded from her observations by Sally Douglas



Complementary Colors, Fig. 3 Afterimage colors as recorded by individual students at Curtin University of Technology

Colors That Are Least Like Each Other

Wilhelm Ostwald points out how “if one moves away from a given hue in the hue circle, the colors become increasingly dissimilar” [12, p. 33]. Since the hue circle is continuous, there will come a point where this dissimilarity is at its greatest, and beyond which there will be a progressive return to similarity. For Ostwald, “there exists for every hue in the hue circle another that is most different from it. This relationship is mutual. The entire hue circle is filled with such pairs of contrasting colors, which shall be called complementary colors” [12, p. 34]. However, the determination of least similarity must depend on the judgment of observers, and such judgments are likely to vary. Ostwald preferred precision. Accordingly he appeals to mixture and proposes an alternative definition: “Complementary colors are colors which in an optical mixture yield a neutral gray” [12, p. 35].

Mixture to a Neutral

When discussing mixture, it is important to distinguish between the paints, inks, or lights that are being mixed, on the one hand, and the appearance of those paints, inks, or lights and the resultant mixture, on the other hand. If a paint that appears

red is mixed with a paint that appears blue, the mixture will appear purple, more bluish, or more reddish depending on how much of each paint is in the mixture. There are different ways of mixing: subtractive, additive, and partitive. The results of one way of mixing are not always a guide to the likely results of the other ways.

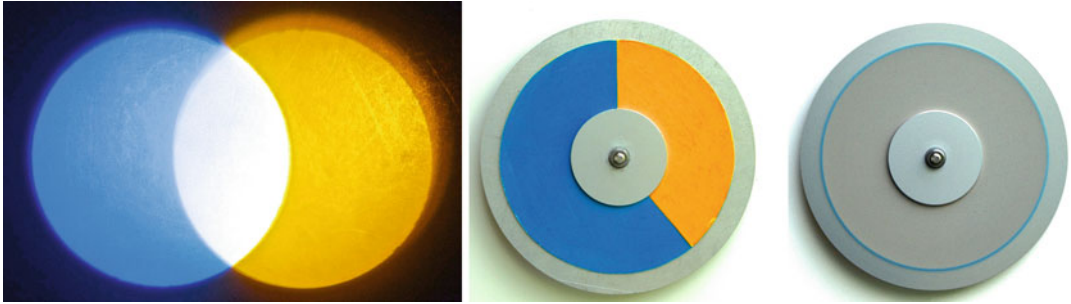
Subtractive Mixture

Subtractive mixture occurs when paints are mixed in the palette, when inks are premixed before being used on the press, and when transparent inks or paints are applied in layers one over the other. Subtractive mixture also occurs when two differently colored filters are placed in front of a single light source. Experiment with different paint combinations can lead to the discovery of a pair of paints that can be mixed to a dark gray as shown in Fig. 4.

Moses Harris claims that if such colors (i.e., paints) are mixed and are “possest of all their powers, they then compose a deep black” [8, p. 7]. But he concedes that no pigments that are generally available, like those used for Fig. 4, have such “powers.” Instead of deep black, the result here is a neutral dark gray.



Complementary Colors, Fig. 4 Dark gray in the middle of a sequence of colors mixed from varying amounts of two paints which appear blue and orange when used straight from the tube



Complementary Colors, Fig. 5 Similar colors in filtered lights and on painted paper discs can mix additively and partitively to a neutral – white or gray

Additive and Partitive Mixture

Additive mixture occurs when beams projected from two filtered light sources overlap on a screen. Partitive mixture occurs when a disc with segments painted in different colors is spun at high speed. The results of additive and partitive mixture are shown in Fig. 5.

Partitive mixture was the method used by Ostwald to establish complementary pairs which he then placed opposite to each other on his circle. Ostwald's circle has yellow opposite blue, as in Fig. 5, where Harris has orange as in Fig. 4.

Possibilities from Physics

In the introduction to his book *The Principles of Harmony and Contrast of Colors and Their Application to the Arts*, Michel-Eugène Chevreul suggests a precise definition: “if we re-united the total quantity of the coloured light *absorbed* by a coloured body, to the total quantity of coloured light *reflected* by it, we should reproduce white light: for it is this relation that two differently coloured lights, taken in given proportions, have of reproducing white light, that we express by the terms *Coloured lights complementary to each*

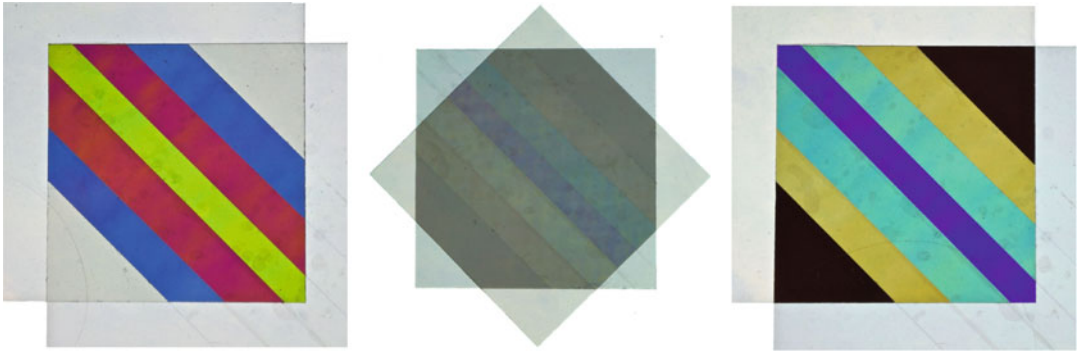
other, or complementary colours” [13, p. 54]. From this, one can imagine a spectral reflectance curve which would serve as a kind of template for a second curve – where one curve had peaks, the other would have valleys. This may be a theoretical ideal; it is doubtful whether two surfaces could be found with such perfectly matched reflectance curves.

Newton's Rings

Color relationships that could be regarded as complementary can be seen in a close examination of Newton's rings. Isaac Newton describes the phenomenon that bears his name in his book *Opticks*. He found that “By looking through . . . contiguous object glasses . . . that the interjacent air exhibited rings of colours, as well by transmitting light as by reflecting it. . . Comparing the coloured rings made by reflection, with those made by transmission . . . I found that white was opposed to black, red to blue, yellow to violet and green to a compound of red and violet” [14, book 2, part I, observ. 9].

Polarized Light and Cellophane Tape

Polarized light can also reveal striking color contrasts when cellophane tape is sandwiched



Complementary Colors, Fig. 6 Colors revealed when cellophane tape is sandwiched between two polarizing filters. The axes of the filters are parallel on the left, at 45° in the center, and at 90° on the right

between two polarizing filters. The filters only transmit light that is vibrating in one plane which means that each filter has what might be called an axis. When two filters are superimposed, they will transmit more or less light depending on the relationship between the axes. If the axes are at right angles to each other, no light is transmitted. The cellophane tape has the effect of modifying the light as it passes through so that certain wavelengths, visible as colors, are transmitted even when the axes are at right angles. This is demonstrated in Fig. 6.

The colors that are seen in the stripes depend on the relationship between the axes and on how many layers of tape there are – in this case, one, two, and three layers, the three-layer stripes being the ones in the center. The colors change radically as one filter is rotated in relation to the other. The colors on the left in Fig. 6 could be described as being complementary to those in the corresponding positions on the right. A more detailed, but simple, explanation of this effect is provided for the “Polarized Light Mosaic” which is included in the *Science Snackbook* compiled by the Exploratorium Teacher Institute [15, p. 78].

Contradictory Complementaries

It is clear from the above that there is no single complementary color for each hue in the color circle. A study to establish complementaries by different definitions was carried out in 1981.

Gouache paints, manufactured by Winsor and Newton (WN) and Pelikan (P), were used for this study. The results are illustrated in Fig. 7.

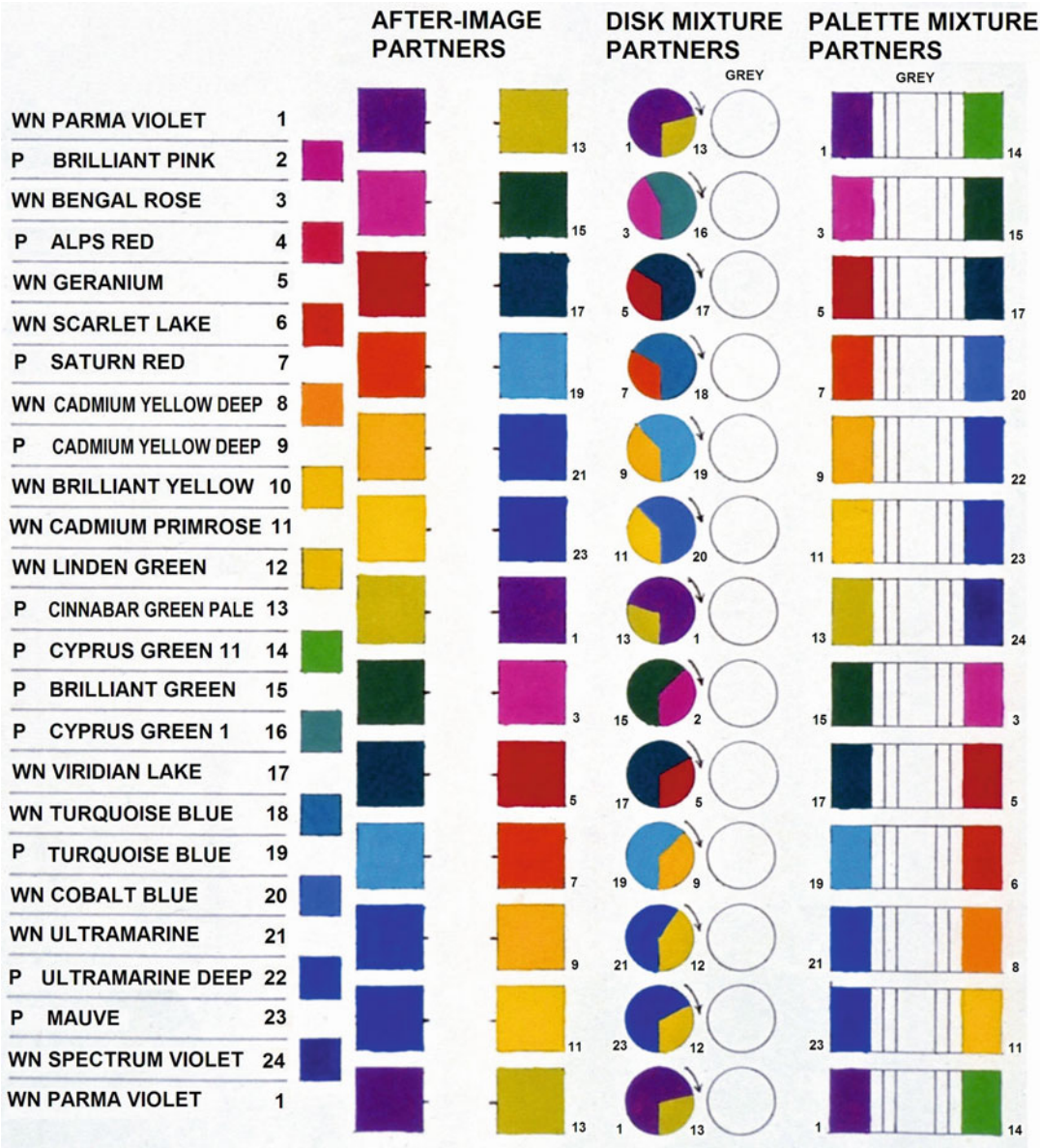
While carrying out the study, it was assumed that two colors can be each other’s afterimage. Given the findings of Sally Douglas, shown above in Fig. 2, such a color circle cannot be constructed except to suggest that opposite colors are close to being each other’s afterimage. The afterimage of unique blue may be a yellow orange, but the afterimage of that yellow orange may be a blue that is very slightly reddish.

An Elastic Color Circle

The issue of contradictory complementaries is discussed in the entry on the color circle in this encyclopedia. An elastic color circle is proposed as a way of illustrating how the number of steps between the unique hues would need to be increased or decreased to bring differently defined complementary pairs opposite to each other. The illustration from the color circle entry is repeated here as Fig. 8.

Complementary Colors as a Guide to Mixing

Complementary colors can be used as a guide to mixing paints, inks, or lights. Given that the results of subtractive and additive/partitive mixture are not the same, the elastic color circle can help to clarify the situation. Figure 9 shows an

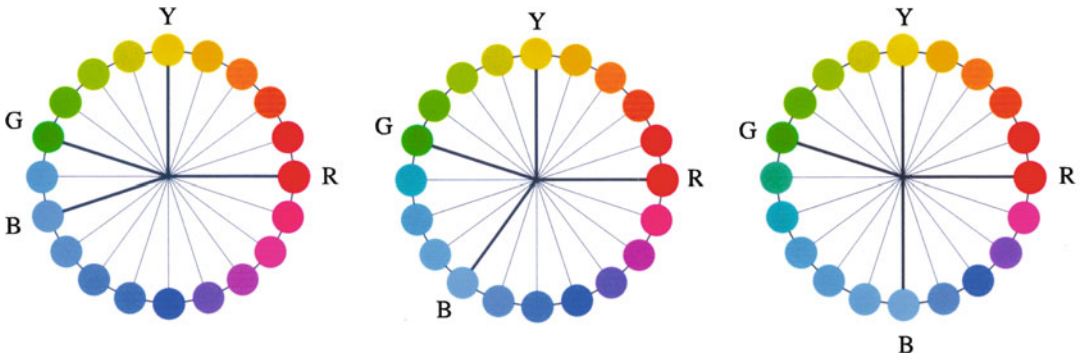


Complementary Colors, Fig. 7 Chart showing the results of a study to find complementary color pairs (partners) according to different definitions

extreme example of the way in which the same two paints can be used to produce radically different results from subtractive and partitive mixture. Here the blue and the yellow are more reddish than the blue and yellow in Fig. 5. The result of subtractive mixture is a dull green, but on the spinning disc, the result is a dull pink.

Lines connecting that blue and yellow pass through green on the subtractive circle, but through pink (red) on the additive/partitive circle, as shown in Fig. 10.

Painters are often advised to use complementary colors, rather than black, white, or gray, to modify the appearance of colors in their painting

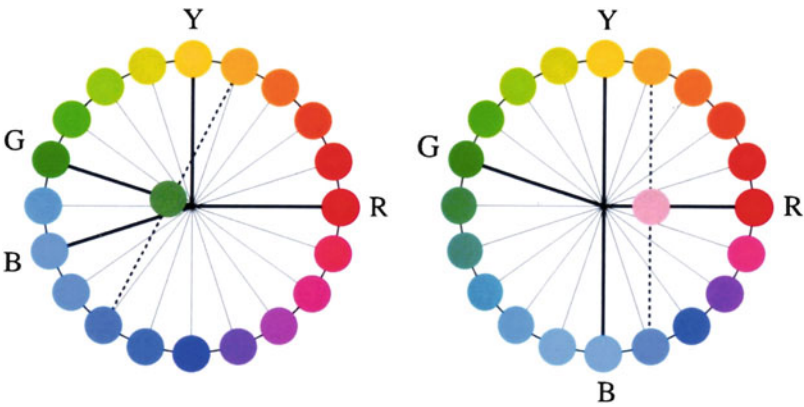


Complementary Colors, Fig. 8 Color circles based on subtractive mixture complementaries (left), afterimage complementaries (center), and additive mixture complementaries (right). The color circles illustrated here were developed from the results of the study shown above in Fig. 7



Complementary Colors, Fig. 9 The same two paints mix subtractively to a dull green and partitively to dull pink

Complementary Colors, Fig. 10 Dotted lines trace the possible results from subtractive (left) and additive/partitive mixtures (right) of a reddish blue and a reddish yellow



that they feel are too vivid. Black, white, and gray paints are felt to have a deadening effect.

Complementary Colors as a Guide to Color Harmony

Complementary colors underpin many theories of color harmony, notably the theories of

Johannes Itten, author of what may still be the most influential and widely used books on color theory for students of art and design today [16, 17]. Itten argues that “the concept of color harmony should be removed from the realm of subjective attitude into that of objective principle” [17, p. 19]. He is unequivocal when he claims that “we can make the general statement that all complementary pairs ... are harmonious” [17,

p. 21]. And he supports this claim by offering three objective means of establishing complementary relationships: “the after-image always turns out to be of the complementary color . . . colors (lights) are harmonious if they mix to give white . . . colors (paints or pigments) are . . . harmonious if their mixture yields a neutral gray” [17, pp. 19–20]. These three definitions cannot coexist in a single color circle. There would need to be separate circles, one for each definition, if complementary colors are to be opposite to one another, and that would mean three possible frameworks for harmonious color combinations. If blue is to be one of the colors in a harmonious two-color combination, the other color would be orange, orange yellow, or yellow, depending on which circle was used. These combinations are shown in Fig. 11.

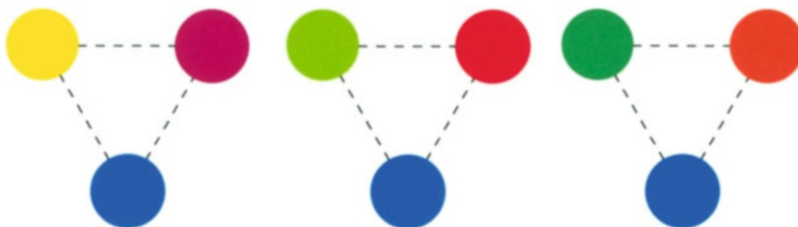
According to Itten, any regular geometric figure can be placed inside the color circle, and its vertices will connect with colors that will form a harmonious combination. An equilateral triangle can be used to find three-color combinations, or triads. Again, with blue as a constant, three rather different combinations would be found in the three circles. These are shown in Fig. 12.

Choosing a Single Definition as the Key to Harmony

Robert Hirschler discusses the potential for confusion arising from different ways of defining complementary colors. Hirschler quotes David MacAdam who was concerned with problems of measurement: “only that definition which states that the optically additive mixture of two complementary colors must match some arbitrarily assigned ‘neutral’ stimulus is sufficiently specific” [18, p. 1]. This way of establishing complementaries appealed to Ostwald and he used it as the basis for his theories of color harmony. And it is certainly true that optical mixture with spinning discs (partitive mixture) was the easiest way to establish precise complementary relationships during the study described above. A large range of vivid and consistent colors can be produced when gouache paints are used straight from the tube. Paper discs were painted and interleaved in very many combinations of two. The proportions were adjusted until particular combinations in particular proportions were found which would spin to appear neutral gray. These are illustrated in Fig. 7. However, it does not follow that the degree of precision obtainable with this process means that this must be the only “correct” way of establishing complementary color pairs which would be, by definition,



Complementary Colors, Fig. 11 Blue with three different complementary colors for three different definitions: (left to right) subtractive mixture, afterimages, and additive/partitive mixture



Complementary Colors, Fig. 12 Blue as one of three colors in combinations found by an equilateral triangle placed inside each of three color circles where

complementaries are established by, from left to right, subtractive mixture, afterimages, and additive/partitive mixture

harmonious. For Goethe, the key to harmony was afterimages, “a natural phenomenon immediately applicable to aesthetic purposes” [9, p. 320]. Afterimages, as subjective experiences, are not so easy to measure but they involve the observer more intimately, and the experience of an afterimage suggests that the visual system is seeking some form of completeness. Chevreul imagined two spectral reflectance curves which would produce white light when combined. A color stimulus and its afterimage could be the perceptual equivalent.

Complementary Colors as a Fuzzy Concept

Artists and designers, who are looking for some formula that will lead to harmonious color combinations, may be perplexed to find that the color circle is unstable as it adjusts to comply with different definitions of complementary colors. Rather than worry about which definition, and its corresponding color circle, is “correct,” it might be more fruitful to accept complementary colors as a fuzzy concept and make decisions based on personal judgment. After studying the color combinations illustrated in Figs. 11 and 12, a choice can be made. If one combination seems more satisfying and harmonious than the other two, the corresponding color circle could be adopted and Itten’s ideas applied. This could give the artist or designer a feeling of ownership while still providing a sense of security with a clear framework for developing harmonious color combinations. Alternatively, a more flexible approach could be taken. If information about the likely results of mixing is needed, then reference could be made to the subtractive or additive circles as appropriate. But as a framework for developing harmonious color combinations, the color circle could be regarded as offering a choice within limits. While a given color may have more than one complementary, as different definitions are applied, those complementary colors will not be radically different in hue. Even in the most extreme example illustrated here, the alternative complementaries for blue, there is not a large spread between yellow and orange. This range can be considered as offering a choice with the

decision being left to the judgment of the artist or designer. It is worth noting that Alberti and Leonardo both nominated more than one color as being “sympathetic” or “contrary.” Alberti mentions red in relation to blue and green and Leonardo lists red or purple or mauve as going well with green.

This more flexible attitude to the color circle is recommended by Deryck Healey. Having introduced a 12-hue color circle based on subtractive mixture, with the printers’ cyan, magenta, and yellow as primaries (the so-called process inks), Healey illustrates combinations that deviate somewhat from strict adherence to Itten’s rules. He advises the reader to “study these examples of harmony and contrast. Few are strictly precise triads, split complementaries and so on; rather they illustrate how such themes may be subtly interpreted and still retain the desired characteristics of their general category” [19, p. 40].

Conclusion

Complementary colors, as defined by mixture to a neutral or by afterimages, provide reference points for the relationships between hues in a color circle. When the relative positions of hues are determined by mixture, the different results from subtractive and additive/partitive mixtures can be recognized and the appropriate circle used as a guide. More problematic is the role of complementary colors in theories of color harmony. Different definitions lead to different relationships between the hues in the color circle. This could lead one to question the theories or abandon them altogether. Nevertheless, the theories are interesting and can be helpful, especially for those who lack confidence in their own intuition. If a choice has to be made between the different circles, there are arguments in favor of objectivity and subjectivity. On the side of objectivity, there is the circle based on additive/partitive mixture where the color pairs are most readily measurable. This has the support of MacAdam and Ostwald. On the side of subjectivity is the circle based on afterimages which relates more directly to the personal experience of color and the judgments of beauty and harmony as made by the observer.

This has the support of Goethe. In the end, it is up to the artist or designer to choose a single circle or to take a more flexible approach as recommended by Healey. An advantage of the elastic color circle is that it allows one to stretch the rules without breaking them.

Cross-References

- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- ▶ [Color Circle](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Harmony](#)
- ▶ [Color Mixture](#)
- ▶ [Goethe, Johann Wolfgang von](#)
- ▶ [Itten, Johannes](#)
- ▶ [MacAdam, David L.](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Palette](#)
- ▶ [Pigments](#)
- ▶ [Unique Hues](#)

References

1. Thompson, D. (ed.): *The Concise Oxford Dictionary*. Oxford University Press, Oxford (1995)
2. Sadie, S. (ed.): *The New Grove Dictionary of Music and Musicians*. Macmillan, London (1980)
3. Green-Armytage, P.: Complementary colours – Description or evaluation? In: Sivik, L. (ed.) *Colour and Psychology: From AIC Interim Meeting 96, Colour Report F 50*, pp. 205–209. Scandinavian Colour Institute, Stockholm (1996)
4. Alberti, L.B.: *On Painting*. Penguin, London (1991). 1435.
5. da Vinci, L.: *Leonardo on Painting*. Yale University Press, New Haven (1989 [1518])
6. Kemp, M.: Note. In: *Leonardo on Painting*. Yale University Press, New Haven (1989)
7. Kemp, M.: *The Science of Art*. Yale University Press, New Haven (1990)
8. Harris, M.: *The Natural System of Colours*. Whitney Library of Design, New York (1963 [1776])
9. von Goethe, J.W.: *Theory of Colours*. MIT Press, Cambridge, MA (1970 [1810]).
10. Cabot Hale, N.: *Abstraction in Art and Nature*. General Publishing Company, Toronto (1993 [1972]).
11. Bagley, M.-O.: Color in design education. In: Arnkil, H., Hämäläinen, E. (eds.) *Aspects of Colour*. University of Art and Design, Helsinki (1995)
12. Ostwald, W.: *The Color Primer*. Van Nostrand Reinhold, New York (1969 [1916]).
13. Chevreul, M.-E.: *The Principles of Harmony and Contrast of Colors and their Application to the Arts*. Reinhold Publishing, New York (1967 [1839]).
14. Newton, I.: *Opticks. Impression Anastaltique, Culture et Civilisation*, Brussels (1996 [1704]).
15. Institute Exploratorium Teacher: *Science Snackbook*. The Exploratorium, San Francisco (1991)
16. Itten, J.: *The Art of Color*. Van Nostrand Reinhold, New York (1961)
17. Itten, J.: *The Elements of Color*. Van Nostrand Reinhold, New York (1970)
18. Hirschler, R.: Teaching colour wheels and complementary colours. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) *Colour – Effects and Affects, Interim Meeting of the AIC, Proceedings, paper 98*. Scandinavian Colour Institute, Stockholm (2008)
19. Healey, D.: *Living with Colour*. Macmillan, London (1982)

Compositing and Chroma Keying

Jorge Lopez-Moreno

GMRV Group, Universidad Rey Juan Carlos,
Madrid, Spain

Synonyms

[Blending](#); [Color-separation overlay](#); [Deep compositing](#); [Layering](#); [Matting](#)

Definition

Compositing is the act of combining two images or video sequences, producing a new image or video sequence. There are several techniques to composite two images (or video frames): alpha matte, gradient-domain blending, deep compositing, multiply, overlay and screen modes, etc.

Chroma keying is the name of one of such techniques which uses color hues (chroma) to guide the compositing process. A portion of the

target image is selected (based on a color or a range of colors) and substituted by the image to be inserted. This technique is widely adopted in video editing and postproduction.

Origin of the Terms

Traditionally, both in TV and films, there have been four basic compositing techniques: matting, physical compositing, background projection, and multiple exposure.

- **Matting** is currently the most widespread technique and corresponds to the general definition given above. Digital compositing relies entirely on variations of this technique: *chroma keying* is an example of matting.
- **Physical compositing**: When capturing the background image, an object is physically introduced into the scene to be composited as foreground image. For instance, a *glass shot* consists of recording a scene through a transparent glass where some elements (or most of it) are painted on the glass (some background buildings, for instance). The area of the frame where the action happens is left clear.

- **Multiple exposure**: One of the earliest compositing techniques ever developed, achieved by recording multiple times with the same film, but exposing a different part each time with the help of a mask over the lens. Georges Méliès, pioneer of visual effects, used it to obtain multiple copies of himself in the film the *One-man band* (Fig. 1).
- **Background projection**: This technique, currently fallen in disuse, is based on projecting the desired video or image onto a background screen with the foreground elements (actors, objects) between the camera and this screen. The development of digital compositing and its own complexities (synchronization issues, illumination constraints) rendered this method obsolete.

Techniques

This section describes some of the standard techniques in image compositing together with the latest advances in the last decade. For further exploration of digital compositing, visual effects, and specific examples in the film industry, please refer to the work of Ron Brinkmann [1].

Compositing and Chroma Keying,

Fig. 1 One of the first examples of compositing in films. A frame from George Méliès *One-man band* movie (1900)



Chroma Keying

This method renders a range of colors in the foreground image as transparent, revealing the image below. Blue and green are the most used color in films, videogames, and TV due to their hue distance from the human skin tone. The main disadvantage of chroma keying is that it requires a lighting set with a (sometimes rather large) chroma screen which has to be as evenly illuminated as possible in order to minimize the range of color variations (noise) in the background (Fig. 2). A secondary issue is that the object (or person) to be inserted cannot have any of the hue values used for chroma keying (e.g., a green dress).

Green is generally preferred over blue due to the lower energy required to produce an even illumination over the chroma screen. However, sometimes blue is used whenever there is a risk of green tones appearing in the foreground layer (e.g., outdoor scenes with vegetation which produce green interreflections). In computer graphics, chroma keying is usually obtained as a function of RGB values which measures the differences from the range of colors used for chroma keying (which is analogous to finding the distance to a closed 3D surface in space color):

$$\begin{aligned}\alpha(p) &= f(R_0, G_0, B_0) \\ &= d(R_0, R_{ck}) + d(G_0, G_{ck}) + d(B_0, B_{ck})\end{aligned}$$

where d is a distance function (e.g., *Euclidean distance*) and both (R_0, G_0, B_0) and (R_{ck}, G_{ck}, B_{ck}) are the red, green, and blue channel values, respectively, of pixel p in the image and the color used for chroma keying.

Alpha Blending

Compositing is based on the information stored in the alpha channel of the inserted image. Introduced for the first time in the image editor *Paint3* in 1970, the *alpha channel* [2, 3] stores a value between 0 and 1, 0 being a pixel which is completely transparent and 1 completely opaque. This channel in a 2D image is a grayscale image by itself called *matte*, which can be visualized and edited. Multiple file formats support alpha-extended data (like RGBA): PNG, TIFF, TGA, SGV, and OpenEXR.

Although binary alpha masking (exclusively 0 or 1) is generally suitable for several compositing scenarios, complex visual phenomena such as transparency or translucency require a subtle gradation of alpha values (e.g., separating a dog from its background requires dealing with strands of hair which capture and scatter the environment light).

Still an open problem, researchers have developed complex methods to obtain these subtle matte masks with the minimal user input. For instance, Levin et al. [4] propose a method called *spectral matting*, which finds clusters of pixels

Compositing and Chroma Keying,

Fig. 2 Example of chroma key compositing (*green screen*). The actress is overlaid with a synthetic background in real time on camera. (Source: Deviantart artist: AngryDogDesigns)



with affinity properties (such as X, Y, and/or RGB distances) by relying on spectral analysis (with matte Laplacian) to evaluate automatically the quality of a matte without explicitly estimating the foreground and background colors. This method requires less user input than most approaches (four strokes on average) in order to distinguish the foreground from the background.

Blend Modes

The majority of the image compositing software includes multiple blending modes, that is, different functions to mix each pair of pixels from two overlaid images depending on their RGB values and the layering order. There are several well-known modes [5]: multiply, screen, overlay, soft light, hard light, dissolve, color dodge, addition, subtraction, darken, lighten, etc. Some of them might differ in implementation and formulation, but some of the most established and agreed modes are described here:

- **Multiply:** Both pixel values are multiplied. The result tends to be darker than any of the original images: dark (black) values are preserved, while white values have no effect in the final composition. This mode is especially useful when combining black and white line drawings with color images.
- **Screen:** Considered as the opposite of the previous mode. Both layers are inverted and multiplied. The final result is inverted again. It tends to produce brighter images than the original, as bright values are preserved and black pixels have no effect. If a denotes the

background pixel and b the foreground pixel, the following equation shows how the screen blend mode is applied:

$$f(a, b) = 1 - (1 - a)(1 - b)$$

- **Overlay:** A hybrid mode based on a combination of multiply and screen modes, guided by the value of the background layer. Foreground colors overlay the background while preserving its highlights and shadows:

$$f(a, b) = \begin{cases} 2ab & \forall a \leq 0.5 \\ 1 - 2(1 - a)(1 - b) & \forall a > 0.5 \end{cases}$$

Gradient-Domain Compositing

Gradient-domain techniques aim to merge two images, making the boundary between them imperceptible. They rely on image gradients (differences of pixel values instead of absolute values), looking for the composite that would produce the smoothest fuse of the gradient field (see Fig. 3). These approaches were introduced by Perez et al. in 2003 [6], being now the standard in commercial software (like the *healing brush tool* in Adobe Photoshop).

The latest research advances combine gradient domain with multi-scale methods and visual transfer to produce compositions which mimic even the structural noise at different spatial levels [7].

Deep Compositing

Deep compositing techniques, in addition to the usual color and opacity channels, take into account depth information stored at each pixel along the



Compositing and Chroma Keying, Fig. 3 Example of image compositing based on gradient domain and proximity matching (Photoshop). An area from the input image (left) is selected with a loose stroke (middle), from which a

radius is derived to analyze gradient and color affinity. The result is shown in the right image. Some artifacts are still noticeable and additional user input might be required

z-axis, perpendicular to the image plane (see Fig. 4). In opposition to traditional compositing, which arranges 2D layers in 3D with a single depth value for each element, deep compositing stores a range of depth values for each object, extending the gamut of editing possibilities. For instance, if the compositing artist aims to integrate an actor with a 3D rendered column of billowing smoke, without 3D deep data in the smoke element, the actor will appear to be in front or behind the column. However, with varying density values in the Z-axis, the actor can be placed *inside* the smoke and show both correct occlusion and partial visibility effects.

In order to mix and compose deep images, a proper depth buffer is required. This data is available for traditional 3D rendered graphics, as each pixel has a depth value associated (Z-buffer), whereas for footage obtained from camera, reliable depth values are available only at some arbitrary specified points. This sparsity is due to the limitations of the capturing device (stereo, time-of-flight cameras, etc.). The remaining pixels are then obtained through interpolation. For instance, Richardt et al. [8] incorporate a time of time-of-flight IR camera to a consumer-level video camera in order to capture a rough depth map which is subsequently filtered to obtain a high-quality depth image by means of spatiotemporal denoising and an upsampling scheme (see Fig. 4).

Although one of the first uses was Pixar RenderMan's deep shadow technique [9], this technology has been progressively adopted by the media industry with the upsurge of 3D cinema and TV. Nowadays, most companies like Weta or DreamWorks rely on deep compositing pipelines

and tools such as Nuke (The Foundry) to create their final compositions.

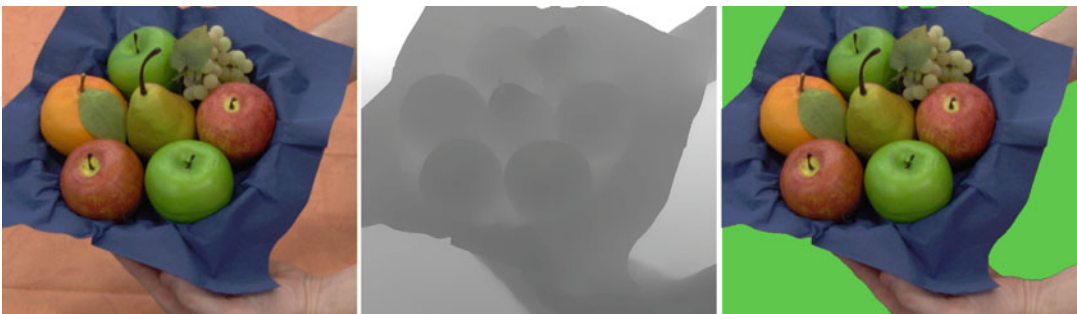
Visual Transfer and Relighting

The concepts of visual transfer and relighting refer to a set of techniques that aim to transfer some visual properties from the background to the image to be inserted. Illumination (shading, shadows, and highlights) is one of the main visual factors to be considered to homogenize a composition. When real objects have to be introduced into CGI environments with known illumination, this illumination is mimicked with actual light sources in a stage. In the early years of cinema, this technique was necessary when performing background projection (e.g., when recording an actor driving a car by night, the lights from cars or street lamps reproduced in the back screen were synchronized with actual lamps illuminating the actor in the studio). More sophisticated systems like Lightstage [10] have been developed to capture an actor performance.

By combining high-speed cameras and structured light sources from many directions, it is possible to create a database of views of the actor under different light environments which can be interpolated to re-render under arbitrary lighting conditions for compositing in any background (see Fig. 5).

Deep Learning-Based Compositing

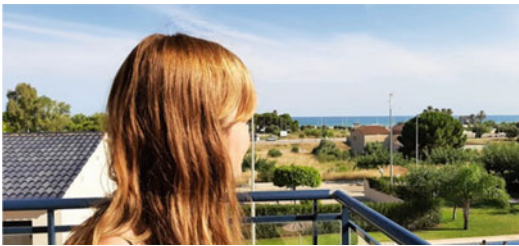
In the last decade, the recent advances in deep learning have made possible accurate relighting



Compositing and Chroma Keying, Fig. 4 Example of deep compositing [8]. The fruit basket in the input image (*left*) is extracted from the background (*right* image, colored in *green*) by using the associated depth range data (*middle* image)

Compositing and Chroma Keying,

Fig. 5 Example of image capture in a Lightstage [9]. The actor is illuminated with an even omnidirectional light in order to capture the base reflectance of his face. In further shots at high speed, the light sources are switched on and off individually to capture the interaction with each light source for future interpolation



Compositing and Chroma Keying, Fig. 6 Example of AI-based background removal method. Left: Original image. Right: Novel composition.

and image composition techniques for static images and video alike. The application of generative neural networks to existing media content of actual people has allowed plausible and realistic reenactment (e.g.: lip sync to novel audio), replacement (e.g.: face swapping), editing, and synthesis creating the phenomenon known as *deepfake*, the generation of believable media by a deep neural network.

This has raised increasing awareness and concern over identity control, spurring the research in the field for deep forgery detection. Most detection techniques are based on either artifact-based detectors (abnormal poses and mannerisms, speech and face consistency), or deep learning classifiers (CNN networks), which have shown certain success but also vulnerabilities to adversarial learning attacks. Readers are referred to the survey by Mirsky and Lee [11] for a

comprehensive description of neural architectures and compositing detectors.

Despite these concerning applications, deep learning has also helped by reducing the need for equipment, scene setups, or even chroma-based backgrounds to perform relighting and background compositing. For instance, the creators of Lightstage, leveraged data obtained with controlled illumination to train a self-supervised system to perform relighting of “in the wild” images that were acquired under arbitrary conditions [12].

Moreover, recent deep learning-based matting methods can extract an accurate alpha channel of entire video sequences just by taking an image of the background without the subject, as shown by the generative adversarial network (GAN) proposed by Segunpta et al. [13]. Some commercial applications of AI-based background removal are available online [14], capable of automatically

extract human shapes from an image, even with challenging components, such as hair, as shown in the left silhouette which is extracted and composed with a new background on the right image at Fig. 6.

Cross-References

► [Global Illumination](#)

References

1. Brinkmann, R.: The Art and Science of Digital Compositing. Morgan Kaufmann, San Francisco (1999)
2. Ray, S. A.: Image Compositing Fundamentals. Microsoft Tech. Memo 4 (1995)
3. Porter, T., Duff, T.: Compositing digital images. In: SIGGRAPH' 84: Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques, pp. 253–259 (1984)
4. Levin, A., Rav-Acha, A., Lischinski, D.: Spectral matting. *IEEE Trans. Pattern. Anal. Mach. Intel.* **30**, 1699–1712 (2008)
5. Grasso, A. (ed.): SVG Compositing Specification. W3C Working Draft (2011)
6. Perez, P., Gangnet, M., Blake, A.: Poisson image editing. *ACM Trans. Graph.* **22**(3), 313–318 (2003)
7. Sunkavalli, K., Johnson, M.K., Matusik, W., Pfister, H.: Multi-scale image harmonization. *ACM Trans. Graph.* **125**, 1–125 (2010). SIGGRAPH, 10
8. Richardt, C., Stoll, C., Dodgson, N., Siedel, H.-P., Theobalt, C.: Coherent spatiotemporal filtering, upsampling, and rendering of RGBZ videos. In: Computer Graphics Forum Proceedings of Eurographics. Eurographics Association, Cagliari (2012)
9. Lokovic, T., Veach, E.: Deep shadow maps. In: ACM SIGGRAPH 2000, pp. 85–392 (2000)
10. Debevec, P.: Virtual cinematography: relighting through computation. *IEEE Comput.* **39**(8), 57–65 (2006)
11. Mirsky, Y., Lee, W.: The creation and detection of deepfakes: a survey. *ACM Comput. Surv.* **54**(1), Article 7 (2021)
12. Sun, T., Barron, T., Tsai, Y., Xu, Z., Yu, X., Fyffe, G., Rhemann, C., Busch, J., Debevec, P., Ramamoorthi, R.: Single image portrait relighting. *ACM Trans. Graph.* **38**(4), Article 79 (2019)
13. Sengupta, S., Jayaram, V., Curless, B., Seitz, S., Kellemacher-Shlizerman, I.: Background Matting: The World is Your Green Screen. *Computer Vision and Pattern Recognition (CVPR)* (2020)
14. www.remove.bg, Developed by Kaleido AI GmbH. Link checked in August 2021

Compton Effect

► [Compton Scattering](#)

Compton Process

► [Compton Scattering](#)

Compton Scattering

P. R. B. Saull

Metrology Research Centre – Ionizing Radiation Standards, National Research Council, Ottawa, ON, Canada

Synonyms

[Compton effect](#); [Compton process](#); [Compton-Debye effect](#); [Incoherent scattering](#)

Definition

Compton scattering is the process whereby a photon (usually a gamma-ray or X-ray) transfers a nonzero fraction of its energy to a charged particle, usually an electron. The particles in the initial and final states are the same (a photon and an electron), making this a scattering process.

For the purposes of this entry, it is assumed that the scatterer is an electron. Typically, this is an atomic electron in a material that has been penetrated by an X-ray or gamma-ray. For lower-energy photons, other processes prevail, such as Rayleigh and Mie scattering or photoelectric absorption.

Photon Interaction Processes

In the visible and UV range, where photon wavelengths are much larger than the atom's size,

interactions tend to be discussed in terms of beams of light impinging on bulk materials using terms such as absorption, transmission, reflection, and refraction. For much higher photon energies in the X-ray and gamma-ray ranges, interactions with individual electrons and nuclei take place. These are typically described in terms of cross sections, although attenuation and absorption coefficients are also widely used for characterizing interaction probabilities in bulk materials. The three main photon interactions for X-ray and gamma-rays are as follows:

- *Photo-electric absorption*, where the photon is completely absorbed and an electron requires enough energy to be ejected by the atom with some kinetic energy
- *Pair-production*, where the photon interacts with the field of the atomic nucleus to produce an electron and a positron (antielectron)
- *Compton (incoherent) scattering*, where the photon scatters off an atomic electron, ejecting it from the atom

This entry covers the third interaction process – Compton scattering.

The Compton scattering process is perhaps most familiar to those working in the field of X-ray and gamma spectrometry, the measurement of gamma and X-ray energies typically in the 1 keV to 3 MeV range, which is the range over which naturally occurring radioactive material (NORM), or natural background radiation, produces photons. Radiation therapists treat cancer patients using electron linear accelerators that produce photon beams in the 6 to 30 MeV range which undergo the Compton process as they deposit their energy in the tumor target. Particle physicists are also very familiar with Compton scattering as it is one of the processes that a high-energy photon, typically many GeV, will undergo as it deposits its energy in a detector's calorimeter through a "shower" process. This starts predominantly as a complex interplay of alternating pair-production and electron/positron *Bremsstrahlung* (radiative photons that in turn undergo pair-production) and eventually gives

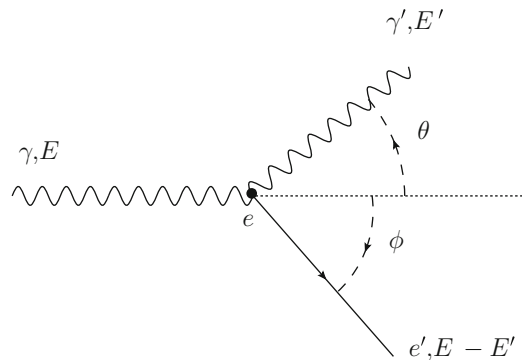
way to the Compton and photo-electric absorption processes in the low-energy tail below a few MeV.

History

In the early twentieth century, X-rays were observed to scatter off materials with a wavelength lower than that of the primary beam, in conflict with the classical wave theory (Thomson scattering) assumed at the time which predicted the wavelength would be unchanged. Arthur Haley Compton, an American physicist, provided an explanation for this observation in 1923 [1] by considering the photon as a particle having quantized energy, as with Einstein's explanation of the photo-electric effect, and assuming a billiard ball type of interaction with the electron that satisfied energy and momentum conservation. For this discovery, Compton was awarded the 1927 Nobel Prize in Physics. The Compton effect was also discovered independently by the Dutch-American physical chemist Peter Debye in early 1923, and the Compton-Debye formula is a kinematic description of this effect.

Kinematics

Figure 1 depicts a photon γ with initial energy $E = h\nu$ impinging upon a stationary electron e of rest mass m_e and scattering (γ') at angle θ with



Compton Scattering, Fig. 1 Compton scattering process, $\gamma e \rightarrow \gamma' e'$

energy E' . The electron e' recoils at angle ϕ with energy $E - E'$.

Based on the energy ranges discussed in section “[Photon Interaction Processes](#),” a 1 MeV photon is somewhat representative of the photon energies to be considered. Assuming half this energy (0.5 MeV) is imparted to the electron through the scattering process, the electron would increase in energy by an amount roughly equal to its rest-mass energy of $m_e c^2 = 0.511$ MeV, making the kinematics of the process relativistic. This is easily seen by comparing the initial and final energy of the electron, E_i and E_f , respectively, using Einstein’s equation for a mass at rest and in motion with additional (kinetic) energy $\delta E = m_e c^2$:

$$\begin{aligned} E_i &= m_e c^2 \\ E_f &= m c^2 = \gamma m_e c^2 = E_i + \delta E = 2m_e c^2 \end{aligned}$$

where $\gamma = \sqrt{1 - \beta^2}$, $\beta = v/c$, and v is the final velocity of the electron in the lab frame. Solving yields $\beta \approx 0.87$. Therefore, relativistic kinematics must be used in the calculation of the Compton process.

In the remainder of this entry the natural units of particle physics are used, $c = \hbar = 1$, such that energy [MeV], momentum [MeV/c], and mass [MeV/ c^2] all have the same units [MeV].

The easiest way to proceed is to use four-momentum notation, $P^\mu = (E, P_x, P_y, P_z)$ with $P^2 = P_\mu P^\mu = E^2 - P_x^2 - P_y^2 - P_z^2 = E^2 - \mathbf{P}^2 = M^2$, and write down the energy-momentum conservation equations. Following Fig. 1,

$$q + p = q' + p' \quad (1)$$

where q and p (q' and p') are the initial (final) four-momenta of the photon and electron, respectively. Dropping the subscript on m_e for convenience,

$$q = (E, 0, 0, E) \quad (2)$$

$$p = (m, 0, 0, 0) \quad (3)$$

$$q' = (E', E' \sin \theta, 0, E' \cos \theta) \quad (4)$$

$$p' = (E'_e, -p'_e \sin \phi, 0, p'_e \cos \phi) \quad (5)$$

$$= (E + m - E', -E' \sin \theta, 0, E - E' \cos \theta) \quad (6)$$

where four-momentum conservation has been used in the last line to eliminate the dependence on variables E_e and p_e . Rearranging, squaring, and making use of the invariant mass relations $q^2 = q'^2 = 0$ and $p^2 = p'^2 = m^2$ yields an expression for the scattered photon energy E' as a function of scattering angle:

$$\begin{aligned} (q - q')^2 &= (p - p')^2 \\ -2qq' &= -2pp' + 2m^2 \\ qq' &= pp' - m^2 \end{aligned}$$

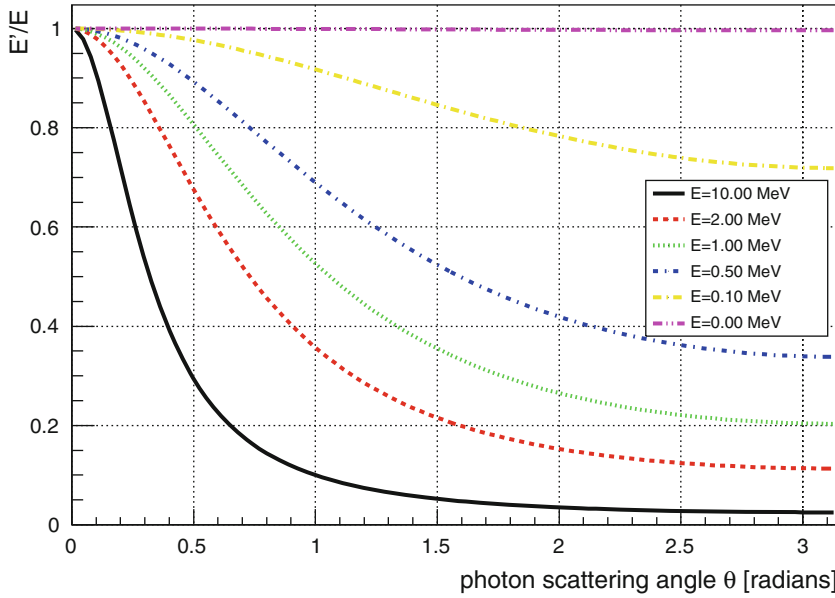
This manipulation has the advantage that the x and y components of the four-momentum fall out in both dot products, leaving an expression easily solved for the ratio of scattered to initial photon energy, E'/E :

$$\begin{aligned} EE' - EE' \cos \theta &= m(E + m - E') - m^2 \\ &= m(E - E') \\ EE'(1 - \cos \theta + m/E) &= mE \\ \frac{E'}{E} &= \frac{1}{E/m(1 - \cos \theta) + 1} = \frac{\lambda}{\lambda'} = \frac{k}{k'} \end{aligned} \quad (7)$$

where use has been made of the relation $E = h\nu = hc/\lambda = \hbar ck$ to express the ratio in terms of wavelength (λ, λ') and wave number (k, k'). Figure 2 shows the fraction of initial energy carried by the scattered photon E'/E as a function of scattering angle θ , for different initial energies. For forward-going scatters, the photon retains most of its energy, as these are glancing collisions with small transverse momentum transfer. All photon-scattering angles are kinematically accessible. In the low-energy limit, Eq. 7 reduces to $E' = E$, independent of angle θ , and consistent with expectations from classical wave theory.

By four-momentum conservation, the scattered electron is given additional kinetic energy $T = E - E'$. Its scattering angle ϕ can be

Scattered photon energy versus angle



Compton Scattering, Fig. 2 Scattered photon energy versus angle

constructed by equating the ratio of transverse to longitudinal p' momentum from Eqs. 5 and 6:

$$\begin{aligned} \tan \phi &= \frac{\sin \theta}{E/E' - \cos \theta} \\ &= \frac{1}{(1 + E/m) \tan \theta/2} \end{aligned} \quad (8)$$

where Eq. 7 has been used to substitute for E/E' . Since θ ranges from 0 to π , the maximum scattering angle of the electron is $\pi/2$, as expected from a billiard-ball type of interaction. From inspection of Eq. 7, the minimum energy a scattered photon can have is at $\theta = \pi$ radians,

$$E'_{\min} = \frac{E}{2E/m + 1} \quad (9)$$

i.e., a full back-scatter collision imparting maximum energy to the electron:

$$T_{\max} = E - E'_{\min} = \frac{2E}{2 + m/E} \quad (10)$$

In fact, since $1 - \cos \theta$ is a slowly varying function of θ near π (less than 10% change from 180° to 145°), the electron energy is close to the maximum for a wide range of back-scattered photon angles, leading to the well-known *Compton edge* phenomenon in gamma spectrometry which results from Compton scattering in a gamma detector. This will be made more evident when the cross section is considered in the next section.

Cross Section

Compton also gave an approximate expression for the cross section for photon scattering. The full expression was calculated in 1929 by Klein and Nishina [2], one of the first successful applications of the new Dirac wave theory. The differential cross section (for unpolarized photons) for the production of a scattered photon into solid angle $d\Omega = \sin \theta d\theta d\varphi$ is given by [3]:

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{\alpha^2}{2m^2} \left(\frac{k'}{k}\right)^2 \left(\frac{k}{k'} + \frac{k'}{k} - \sin^2\theta\right) \\ &= \frac{\alpha^2}{2m^2} \left(\frac{1}{\eta(1 - \cos\theta) + 1}\right)^2 \left(\frac{1}{\eta(1 - \cos\theta) + 1} \right. \\ &\quad \left. + \eta(1 - \cos\theta) + 1 - \sin^2\theta\right) \end{aligned} \quad (11)$$

where $\eta = E/m$, $\alpha = \frac{e^2}{4\pi}$ is the fine structure constant, and Eq. 7 has been used. This reduces to the Thomson cross section [4] at low energy, $\eta \rightarrow 0$ ($k \rightarrow 0$) [3]:

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{low } E} \rightarrow \frac{\alpha^2}{2m^2} (1 + \cos^2\theta) \quad (12)$$

The cross section is plotted as a function of photon-scattering angle in Fig. 3 for various photon energies.

It is useful to recast the Klein-Nishina cross section in terms of electron kinetic energy, $T = E - E'$, using Eq. 7:

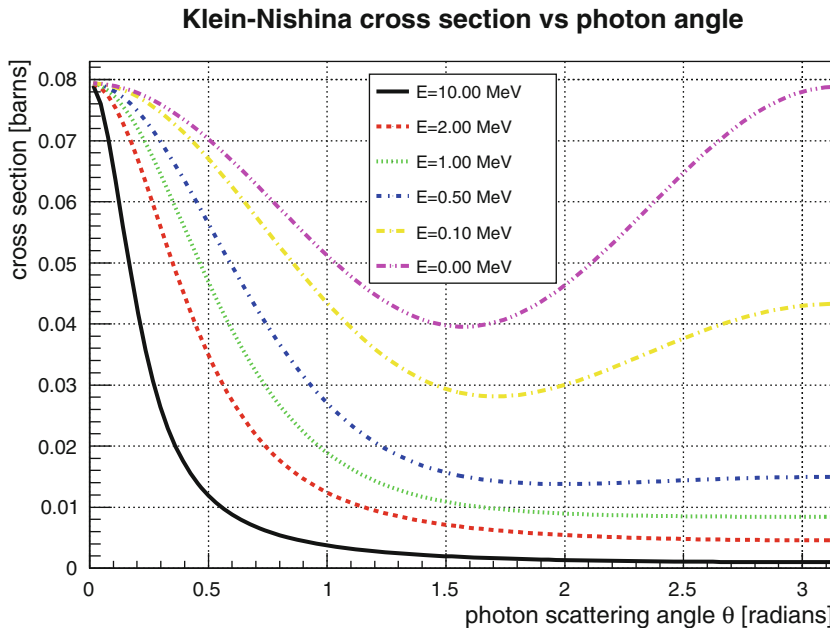
$$\begin{aligned} \frac{d\sigma}{dT} &= \frac{d\sigma}{d\theta} \frac{d\theta}{dT} = 2\pi \sin\theta \frac{d\sigma}{d\Omega} \frac{d\theta}{dT} \\ &= \frac{2\pi}{m} (1 - \cos\theta + 1/\eta)^2 \frac{d\sigma}{d\Omega} \end{aligned} \quad (13)$$

which as a function of T (using Eq. 7) becomes [5]:

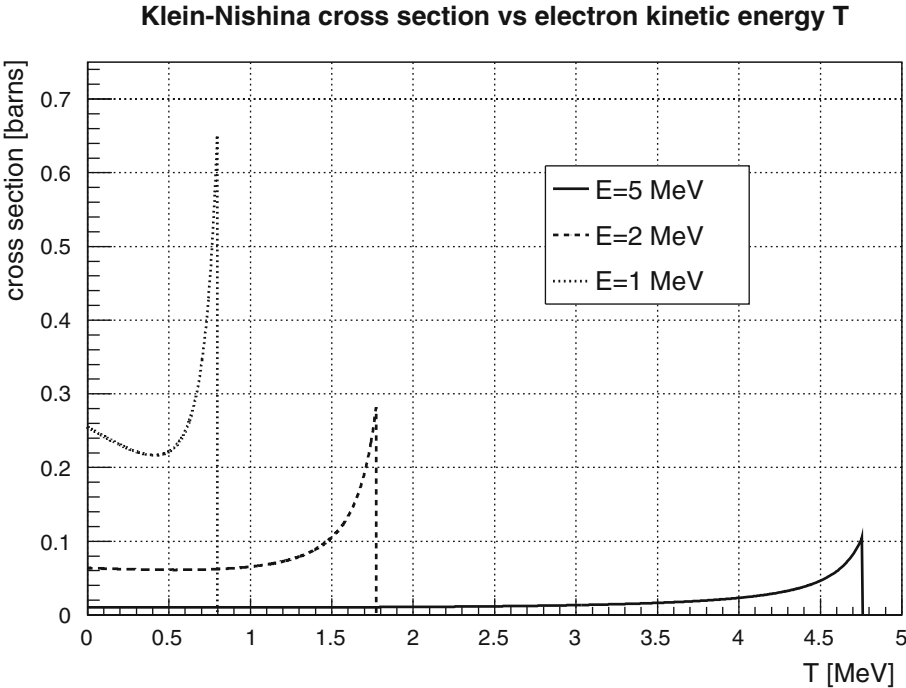
$$\frac{d\sigma}{dT} = \frac{\pi r_e^2}{m\eta^2} \left[2 + \frac{s^2}{\eta^2(1-s)^2} + \frac{s}{1-s}(s - 2/\eta) \right] \quad (14)$$

where $s = T/E$. This is plotted for a few photon energies E in Fig. 4. As motivated in section “Kinematics,” these distributions exhibit a characteristic *Compton edge* at the maximum electron energy. This has good and bad consequences for gamma spectrometry (see section “Applications” below).

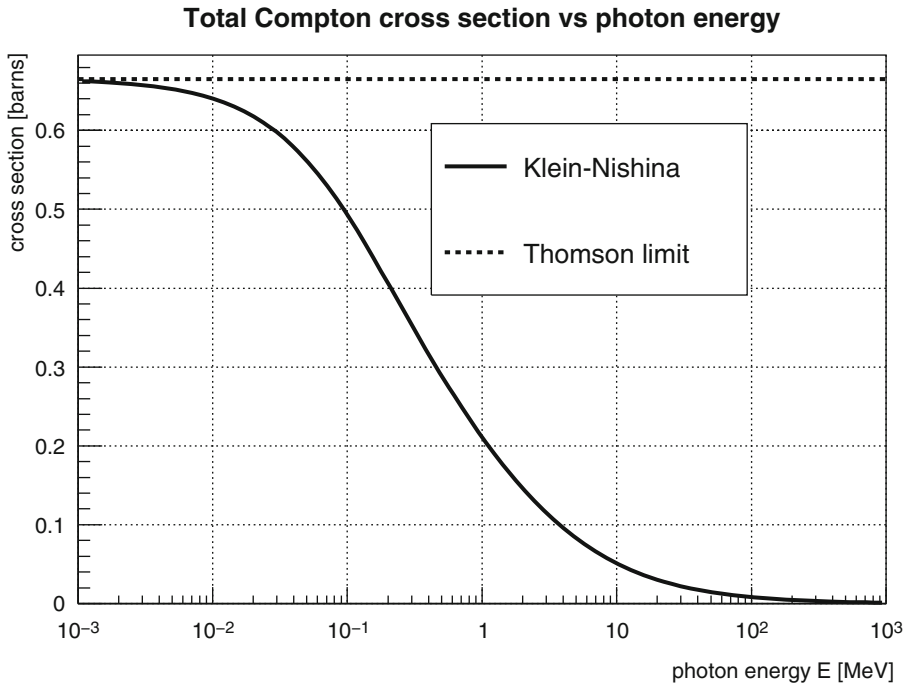
Integrating over solid angle yields the total Klein-Nishina cross section [5] plotted in Fig. 5 versus photon energy. The cross section approaches the classical wave-theory value at the low-energy limit.



Compton Scattering, Fig. 3 Klein-Nishina cross section for a photon scattering off a stationary electron, versus photon scattering angle θ



Compton Scattering, Fig. 4 Cross section as a function of electron energy T



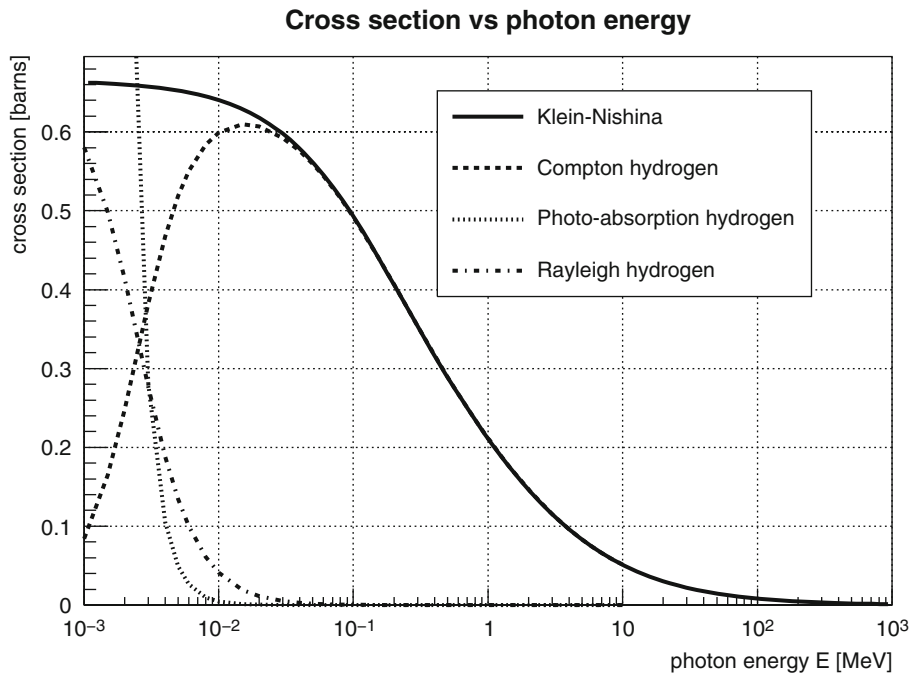
Compton Scattering, Fig. 5 Klein-Nishina total cross section (solid line) as a function of photon energy. Overlaid as a limit (dashed line) is the low-energy Thomson scattering value calculated from classical wave theory

The Klein-Nishina cross section formula was derived for a free electron at rest. For most applications of relevance, electrons are not free but bound by the nucleus of an atom and have momentum. At very low photon energy $E \rightarrow 0$, the photon has long wavelength and essentially interacts with the atom as a whole (different parts of the atom contributing in a coherent way, hence *coherent scattering*), leaving it in an unexcited state with no energy transfer to it. Individual electrons are not resolved, and the probability for *incoherent* (Compton) scattering goes to zero. As the photon energy increases, energy transfer to individual electrons can take place, starting with those least bound in the outer shells. At still higher energy, all electrons become available to the Compton scattering process, appearing as essentially free electrons at rest (compared to the photon energy), and the Klein-Nishina cross section holds more accurately. The overall result of including binding effects is that the Compton cross section is suppressed at low-photon energy with respect to the Klein-Nishina prediction. This is demonstrated in Fig. 6, which shows the Klein-

Nishina cross section (solid line) overlaid with the Compton cross section for hydrogen (long-dash line) using data from the NIST XCOM database [6]. The two cross sections are in good agreement above about $E = 0.1$ MeV, where the photon has a wavelength of $\lambda = hc/E$ of 0.0012 nm. This is small compared to the size of the H atom (the first Bohr radius has value of 0.0529 nm) so that the electron is resolved, appearing essentially free and at rest. At lower photon energy, the curves diverge rapidly as binding effects come into play.

Applications

There are many applications of Compton scattering in a variety of fields including radiobiology, radiotherapy, nuclear medicine, densitometry, nuclear imaging, gamma spectroscopy, material physics, and astrophysics. Only two are covered here in detail: the energy calibration of organic scintillators used in gamma spectrometry; and gamma imaging for medical or safety and security purposes.



Compton Scattering, Fig. 6 Cross sections as a function of photon energy

Gamma Spectrometry

Scintillators are widely used in gamma spectrometry as materials which convert the kinetic energy of ionizing particles into detectable light in the UV or optical range. The scintillating photons are promptly emitted, typically within a *microsecond* of energy deposition, and their number is to first order proportional to the energy deposited by the charged particle. By capturing and directing these photons efficiently to a photodetector (typically a conventional photomultiplier tube or solid-state silicon-photomultiplier device), the energy deposited can be measured as a current pulse with total charge characterizing the energy deposited. For X-rays or gamma-rays, conversion to charged particles through one or more steps of photoelectric absorption, Compton scattering, or pair-production is required for detection. To calibrate such a system for gamma spectrometry, one relies on one or more radioactive sources (e.g., Cs-137) with emission at known characteristic energies (662 keV for Cs-137). Repeated full-energy absorption of these gamma rays leads to “photo-peaks” in the energy spectrum at the characteristic energies, which can then be used to determine the relationship between charge (or pulse height) and energy.

Of course, gamma rays do not always deposit their energy fully – many might scatter once or more and exit the scintillator. Single-scatter interactions lead to the Compton edge structure exhibited in Fig. 4. Generally, such edges are a nuisance for spectrometry as they can make it difficult to isolate the photopeaks for proper measurement, especially for sources with multiple characteristic energies. However, there is an important class of scintillator which relies heavily on the Compton edge for calibration and measurement. These are organic scintillators having low-atomic number for which the probability of a photo absorption is very low or absent entirely depending upon energy (note the cross-section values at 662 keV in Fig. 6). For example, *poly vinyl toluene* (PVT) is an inexpensive plastic scintillator that is robust, nonhygroscopic, and easily tooled and shaped by the end user in the lab. It is the only economical option for deployment in

bulk at border crossings as giant portal monitors past which shipping containers are driven. Unfortunately, the probability for photo-absorption in PVT is so low in most of the energy range of interest (up to 3 MeV) that one must rely solely on the Compton edge to calibrate these detectors and make use of them for gamma identification [7].

Compton Imaging

A second notable application of Compton scattering is in the imaging of gamma-ray fields. Consider a photon detector capable of measuring the energy and position of interactions taking place in it, and for simplicity assume a point source lies in a nearby field of interest. By measuring the energies of the scattered photon and electron, one can make use of Eq. 7 to deduce the angle θ through which the photon must have scattered. Combined with the positions of the scattering point and the detected scattered photon, this yields a handle on the incoming direction of the photon up to a cone of possible angles. (If the direction of the scattered electron can be measured, then the cone reduces to a sector.) Over time, multiple photon interaction “events” yield enough information to deduce where the source must lie (e.g., through simple overlap of the cones back-projected into the field of interest). Extended sources (nonpoint) are more challenging and require more sophisticated algorithms.

Cross-References

- [Light, Electromagnetic Spectrum](#)
- [Photodetector](#)
- [Rayleigh and Mie Scattering](#)

References

1. Compton, A.H.: A quantum theory of the scattering of x-rays by light elements. *Phys. Rev.* **21**(5), 483502 (1923)
2. Klein, O., Nishina, Y.: ber die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen

- Quantendynamik von Dirac. Z. Phys. **52**, 853–868 (1929)
3. Bjorken, J.D., Drell, S.D.: Relativistic Quantum Mechanics. McGraw-Hill, New York (1964)
 4. Thomson, J.J.: Conduction of Electricity through Gases, 2nd edn, pp. 321–331. Cambridge University Press
 5. Leo, W.R.: Techniques for Nuclear and Particle Physics Experiments: A How-to Approach, 2nd revised edn, pp. 56–57. Springer, Berlin (1994)
 6. Berger, M.J., et al.: XCOM: photon cross sections database. <https://doi.org/10.18434/T48G6X>. Last Update to Data Content: November 2010
 7. Siciliano, E.R., et al.: Energy calibration of gamma spectra in plastic scintillators using Compton kinematics. Nucl. Instrum. Methods Phys. Res., Sect. A. **594**, 232243 (2008)

Compton-Debye Effect

- [Compton Scattering](#)

Computation of Colorimetric Stimuli

- [Transformation of Color Matching Functions](#)

Computer Depiction

- [Non-photorealistic Rendering](#)

Computer Modeling

- [Color Category Learning in Naming-Game Simulations](#)

Conceptual Spaces

- [Color Categories and Vagueness](#)

Cone Fundamentals

Andrew Stockman

Department of Visual Neuroscience, UCL
Institute of Ophthalmology, London, UK

Synonyms

[CIE fundamental color matching functions](#); [Cone spectral sensitivities](#)

Definition

The three cone fundamentals are the spectral sensitivities of the long- (L-), middle- (M-), and short- (S-) wavelength cones measured relative to light entering the cornea. They are also the fundamental color matching functions (CMFs), which in colorimetric notation are referred to as $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$. The simple identity between the cone fundamentals of color matching and the cone spectral sensitivities depends on phototransduction and its property of *univariance*. The absorption of a photon produces a photoreceptor response that is independent of photon wavelength, so that all information about wavelength is lost. With one cone type, vision is monochromatic. With three cone types, vision is trichromatic.

Trichromacy means that for observers with normal color vision, the color of a test light of any chromaticity can be matched by superimposing three independent primary lights (with the proviso that one of the primaries sometimes must be added to the test light to complete the match). The amounts of the three primary lights required to match test lights as a function of test wavelength are the three CMFs for those primary lights (usually defined for matches to test lights of equal energy). Because the cone spectral sensitivities overlap, it is not possible for any real light to stimulate just one of them. However, the cone fundamentals are the three CMFs for the three *imaginary* primary lights that would uniquely stimulate individual cone types (i.e., imaginary

lights that produce the three “fundamental” sensations that underlie color vision). The cone fundamental CMFs define all other CMFs and must be a linear transformation of them.

The cone fundamentals can be determined directly by measuring cone spectral sensitivities using “color-deficient” observers lacking one or two cone types and/or special conditions to isolate single-cone responses. They can also be derived by the linear transformation of CMFs measured using real primary lights, but for that the coefficients of the linear transformation must be known.

Current estimates of the cone fundamentals [1–3] use spectral sensitivity measurements to guide the choice of the coefficients of the linear transformation from a set of measured CMFs to the cone fundamental CMFs.

Overview

The spectral properties of the cone fundamentals and that of color matching, in general, are determined principally by the way in which the cone photoreceptors interact with light at the very first stage of vision. They depend, in particular, on photon absorptions by the cone photopigment and on how the probability of photon absorption varies with wavelength. To understand that, we start at the molecular level.

Phototransduction

The cone photopigment molecule is made up of a transmembrane opsin, a G protein-coupled receptor protein, bound to a chromophore, 11-*cis* retinal. The absorption of a photon provides the energy needed to isomerize the chromophore from its 11-*cis* form to its all-*trans* form; this change in shape activates the opsin and triggers the phototransduction cascade and the neural response. The likelihood that a given photon will produce an isomerization depends upon how closely its energy matches the energy required to initiate the isomerization. Crucially, this energy varies with cone type because of differences in key amino acids in those parts of the opsin molecule that surround the chromophore. These amino acids modify the isomerization energy and thus

the spectral sensitivity of the photoreceptor [4]. In most observers with normal color vision, there are four photoreceptor classes: three types of cone photoreceptors, L-, M-, and S-cones, and a single type of rod photoreceptor.

Cone fundamentals, since they are measured behaviorally in terms of energies measured at the cornea, also depend on the absorption of photons by the optical media and on the density of photopigment in the photoreceptor outer segment, both of which vary between observers (see below). Thus, the cone fundamentals are not *precisely* related to the spectral properties of the photopigments (called the absorbance or extinction spectra).

Univariance

The relatively simple relationship between the cone fundamentals and color matching arises because of the way in which cone (and rod) photoreceptors transduce absorbed photons. When a photon is absorbed to initiate the phototransduction cascade, the effect is all or nothing and is consequently independent of photon wavelength. Photoreceptor outputs thus vary *univariantly* according to the number of photons that they absorb [5], as a result of which wavelength and intensity are confounded, and individual photoreceptors are “color blind.” A change in the rate of photon absorption could be due to a variation in light intensity, but equally, it could be due to a variation in wavelength.

With only one cone type, vision is monochromatic and reduced to a single dimension: two lights of any spectral composition can be made to match simply by equating their intensities (a relationship defined by the cone type’s spectral sensitivity). With only two cone types, vision is dichromatic and reduced to two dimensions: lights of any spectral composition can be matched with a mixture of two other lights. Dichromatic human observers fall into three classes, protanopes, deuteranopes, and tritanopes, depending upon whether they lack L-, M-, or S-cones, respectively. Observers with normal color vision have three classes of cone photoreceptor, and their vision is trichromatic. Color vision depends on

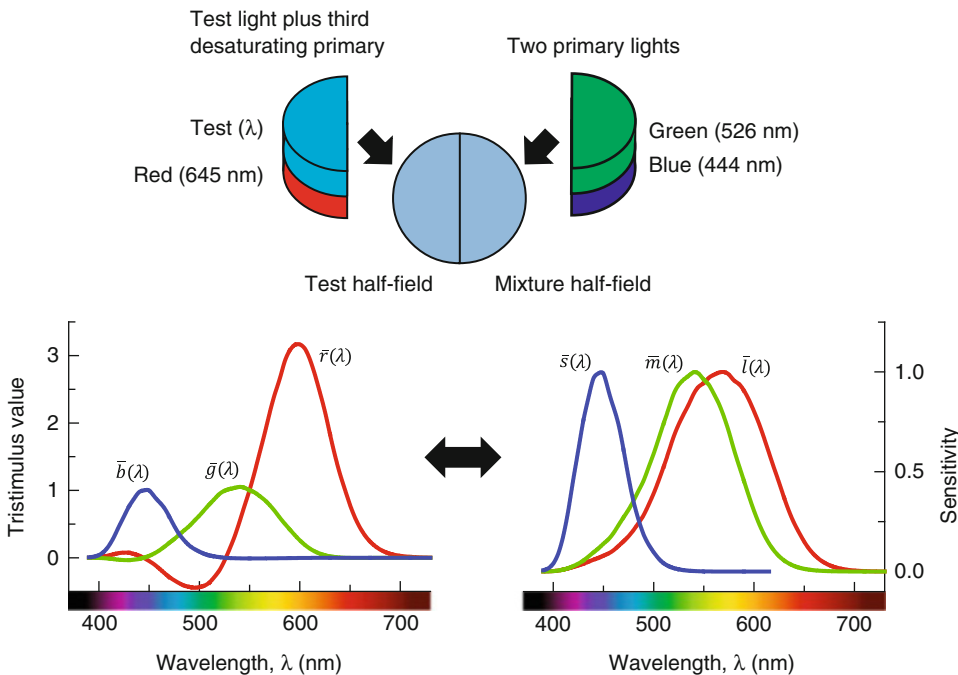
comparing the univariant outputs of different cone types.

Trichromacy

A consequence of trichromacy is that the color of any light can be matched with three specially selected or “independent” primary lights of variable intensity (chosen so that no two will match the third). These primary lights are frequently red (R), green (G), and blue (B), but many other triplets are possible. The upper panel of Fig. 1 shows a typical color matching experiment, in which an observer is presented with a half-field illuminated by a “test” light of variable wavelength λ and a second half-field illuminated by a mixture of red, green, and blue primary lights. At each λ , the observer adjusts the intensities of the three primary lights, so that the test field is

perfectly matched by the mixture of primary lights. The lower left-hand panel of Fig. 1 shows the mean $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ CMFs obtained by Stiles and Burch [6] for primary lights of 645, 526, and 444 nm. Notice that except at the primary wavelengths one of the CMFs is negative. There is no “negative light”; rather, these negative values indicate that the primary in question must be added to the spectral test light to make a match (as illustrated in the panel for the red primary). Real primaries give rise to negative values because real lights do not uniquely stimulate single-cone photoreceptors (see Figs. 2 and 3). The cone fundamental CMFs are always positive.

CMFs, such as the ones shown in the lower-left panel, can be linearly transformed to any other set of real primary lights and to the *fundamental* primaries that are illustrated in the lower



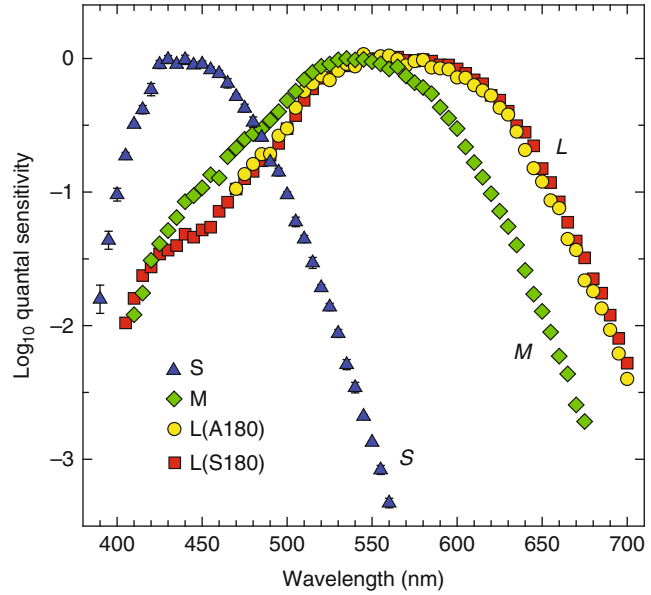
Cone Fundamentals, Fig. 1 A monochromatic test field of wavelength, λ , can be matched by a mixture of red (645 nm), green (526 nm), and blue (444 nm) primary lights, one of which must be added to the test field to complete the match (*upper panel*). The amounts of each of the three primaries required to match monochromatic lights spanning the visible spectrum are the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs (*red, green, and blue lines, respectively*) shown in the lower left-hand panel. These CMFs were measured

using 10-deg diameter targets by Stiles and Burch [6]. A negative sign means that the primary must be added to the target to complete the match. CMFs can be linearly transformed from one set of primaries to another and to the fundamental primaries. Shown in the lower right-hand panel are the 10-deg $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs linearly transformed to give the $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ 10-deg cone fundamental primary CMFs (*red, green, and blue lines, respectively*)

Cone Fundamentals,

Fig. 2 Mean cone spectral sensitivity data [10, 11].

L-cone data from deuteranopes with either the L(S180) (red squares, $n = 17$) or L(A180) (yellow circles, $n = 3$) polymorphism, M-cone data from protanopes (green diamonds, $n = 9$), and S-cone data (blue triangles) from S-cone monochromats ($n = 3$) and below 540 nm (measured under intense long-wavelength adaptation) from normal observers ($n = 5$)



right-hand panel of Fig. 1. The three fundamental primaries (or “Grundempfindungen” – fundamental sensations) are the three imaginary primary lights that would *uniquely* stimulate each of the three cones to yield the L-, M-, and S-cone spectral sensitivity functions (such lights are not physically realizable because of the overlapping spectral sensitivities of the cone photopigments). All other sets of CMFs depend on the fundamental CMFs and should be a linear transformation of them. Note that the fundamental CMFs shown here are comparable to those shown in Fig. 3 but are plotted as linear sensitivities rather than as the more usual logarithmic sensitivities.

The relationship between the fundamental CMFs and a set of real CMFs (obtained using, e.g., red, green, and blue primaries) can be stated formally: When an observer matches the test and mixture fields in a color matching experiment, the two fields cause identical absorptions in each of his or her three cone types. The match, in other words, is a match at the cones. The matched test and mixture fields appear identical to S-cones, to M-cones, and to L-cones. For matched fields, the following relationships apply:

$$\bar{l}_R \bar{r}(\lambda) + \bar{l}_G \bar{g}(\lambda) + \bar{l}_B \bar{b}(\lambda) = \bar{l}(\lambda)$$

$$\bar{m}_R \bar{r}(\lambda) + \bar{m}_G \bar{g}(\lambda) + \bar{m}_B \bar{b}(\lambda) = \bar{m}(\lambda) \quad (1)$$

$$\bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) = \bar{s}(\lambda)$$

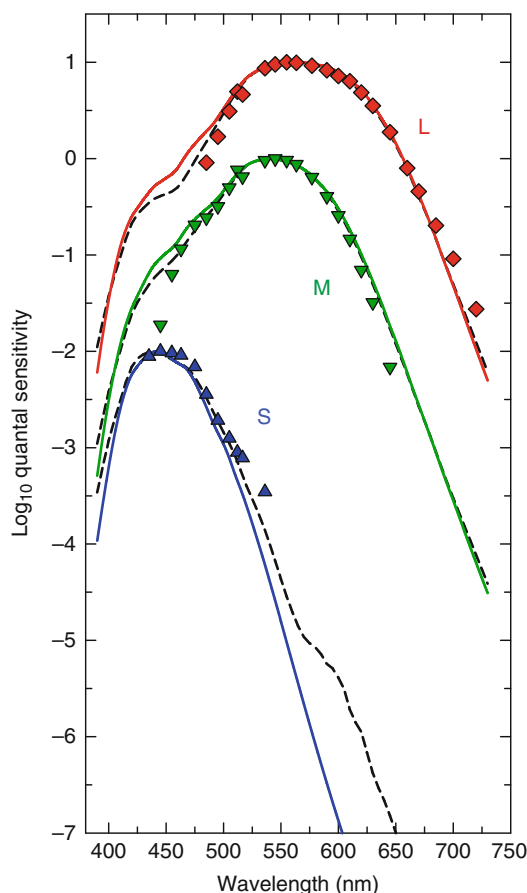
where $\bar{l}_R$, $\bar{l}_G$, and $\bar{l}_B$ are, respectively; the L-cone sensitivities to the R, G, and B primary lights; and similarly, $\bar{m}_R$, $\bar{m}_G$, and $\bar{m}_B$ and $\bar{s}_R$, $\bar{s}_G$, and $\bar{s}_B$ are the analogous M- and S-cone sensitivities. Since the S-cones are insensitive in the long-wavelength (red) part of the spectrum, $\bar{s}_R$ can be assumed to be zero. There are therefore eight unknowns required for the linear transformation:

$$\begin{pmatrix} \bar{l}_R & \bar{l}_G & \bar{l}_B \\ \bar{m}_R & \bar{m}_G & \bar{m}_B \\ 0 & \bar{s}_G & \bar{s}_B \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix}. \quad (2)$$

Since only relative cone spectral sensitivities are required, the eight unknowns reduce to five:

$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix} \quad (3)$$

A definition of the cone fundamental CMFs in



Cone Fundamentals, Fig. 3 Comparisons between estimates of the 2-deg L-, M-, and S-cone fundamentals by Stockman and Sharpe [2] (solid-colored lines), by Smith and Pokorny [1] (dashed lines), and by König and Dieterici [17] (symbols)

terms of real CMFs requires a knowledge of the coefficients of the transformation.

Spectral Sensitivity Measurements

The transformation in Eq. 3 can be estimated by comparing dichromatic and normal color matches [7]. Dichromats confuse pairs of colors that trichromats do not. When these confusions are plotted in a normal chromaticity diagram, they yield characteristic lines of confusion that converge to a different confusion point for each type of dichromat. The three confusion points correspond to the chromaticities of the missing imaginary

fundamental primaries, from which the transformation matrix can be derived.

An alternative, straightforward way of estimating the transformation matrix is to measure the three cone spectral sensitivities directly. This can be achieved using steady or transient chromatic backgrounds to selectively adapt one or two of the cone types to isolate the third [8, 9]. However, cone isolation can more easily be achieved by using chromatic adaptation in observers in dichromats lacking one (or two) of the three cone types. With the S-cones selectively adapted, L- and M-cone spectral sensitivities can be directly measured in deuteranopes without M-cone function and in protanopes without L-cone function. Figure 2 shows the mean spectral sensitivity data obtained from nine protanopes (green diamonds), from seventeen single-gene L(S180) deuteranopes with serine at position 180 of their L-cone photopigment opsin gene (red squares), and from five single-gene L(A180) deuteranopes with alanine at position 180 (orange circles) [2, 10]. L(A180) and L(S180) are two commonly occurring L-cone photopigment polymorphisms in the normal human population that differ in $\lambda_{\max}$ by about 2.5 nm [10]. Figure 2 also shows the mean S-cone spectral sensitivities obtained from three S-cone monochromats, who lack L- and M-cones, and under intense long-wavelength adaptation at wavelengths shorter than 540 nm obtained from five normal subjects [11]. Importantly, thanks to molecular genetics, we can now choose dichromats or monochromats for these experiments whose remaining cone photopigments are normal.

Cone Fundamentals

The spectral sensitivities shown in Fig. 2 were used by Stockman and Sharpe [2] to find the linear combinations of the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs that best fit each measured cone spectral sensitivity, allowing adjustments in the densities of pre-receptor filtering and photopigment optical density in order to account for differences in the mean densities between different populations

and different target sizes. The transformation matrix for the Stockman and Sharpe 10-deg cone fundamentals is

$$\begin{pmatrix} 2.846201 & 11.092490 & 1 \\ 0.168926 & 8.265895 & 1 \\ 0 & 0.010600 & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} \quad (4)$$

where $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs are the Stiles and Burch CMFs measured with a 10-deg diameter test field [6]. The Stockman and Sharpe 2-deg estimates are based on the same transformation but the cone fundamentals have been adjusted to macular and photopigment optical densities appropriate for a 2-deg target field. The 2-deg functions are shown in Fig. 3 as the solid-colored lines. The Stockman and Sharpe 2-deg and 10-deg functions have been adopted by the Commission Internationale de l'Éclairage (CIE) as the 2006 physiologically relevant cone fundamental CMFs [3].

The quality of cone fundamentals depends not only on the correct transformation matrix but also in the CMFs from which they are transformed. The Stiles and Burch 10-deg CMFs [6], which were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm), are probably the most secure and accurate set of existing color matching data. Other CMFs are less secure and typically flawed [12].

Most other cone fundamentals are also given in the form of Eq. 4 but for different underlying CMFs [1, 2, 13–16]. The most widely used have been those by Smith and Pokorny [1]. Their transformation matrix is

$$\begin{pmatrix} 0.15514 & 0.54312 & -0.03286 \\ -0.15514 & 0.45684 & 0.03286 \\ 0 & 0.00801 & 1 \end{pmatrix} \begin{pmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{pmatrix} \quad (5)$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the Judd-Vos-modified 2-deg CMFs [15].

Figure 3 shows the Smith and Pokorny estimates as dashed lines and for historical context the much earlier estimates obtained 125 years ago by

König and Dieterici [17] as symbols. For the L- and M-cone fundamentals, the discrepancies between the more modern fundamentals are mainly at shorter wavelengths; the discrepancies between the S-cone fundamentals are more extensive.

The functions mentioned here can be found at <http://www.cvrl.org>

Other Factors that Influence Cone Fundamentals

Factors other than the properties of the photopigment also affect the cone spectral sensitivities. They include the density of the pigment in the lens that absorbs light mainly of short wavelengths, the density of macular pigment at the fovea, and the axial optical density of the photopigment in the photoreceptor. All three factors exhibit individual differences between observers, and the last two vary with retinal eccentricity. These factors should all be taken into account when trying to predict the spectral sensitivities of an individual from standard functions such as those defined by Eqs. 4 and 5 for a given target size and eccentricity. See, for example, Brainard and Stockman [18] for further details.

Cross-References

- CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments
- CIE Chromaticity Coordinates (xyY)
- CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams
- Deuteranopia
- Evolution of Primate Color Vision
- Protanopia
- Trichromacy

References

1. Smith, V.C., Pokorny, J.: Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vis. Res.* **15**, 161–171 (1975)

2. Stockman, A., Sharpe, L.T.: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vis. Res.* **40**, 1711–1737 (2000)
3. CIE: Fundamental chromaticity diagram with physiological axes – part 1. Technical report 170–1. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
4. Deeb, S.S.: The molecular basis of variation in human color vision. *Clin. Genet.* **67**, 369–377 (2005)
5. Mitchell, D.E., Rushton, W.A.H.: Visual pigments in dichromats. *Vis. Res.* **11**, 1033–1043 (1971)
6. Stiles, W.S., Burch, J.M.: NPL colour-matching investigation: final report (1958). *Opt. Acta.* **6**, 1–26 (1959)
7. Maxwell, J.C.: On the theory of colours in relation to colour-blindness. A letter to Dr. G. Wilson. *Trans. R. Scott. Soc. Arts.* **4**, 394–400 (1856)
8. Stiles, W.S.: *Mechanisms of Colour Vision*. Academic, London (1978)
9. Stockman, A., MacLeod, D.I.A., Vivien, J.A.: Isolation of the middle- and long-wavelength sensitive cones in normal trichromats. *J. Opt. Soc. Am. A* **10**, 2471–2490 (1993.); Published online Epub Dec
10. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., Nathans, J.: Red, green, and red-green hybrid photopigments in the human retina: correlations between deduced protein sequences and psychophysically-measured spectral sensitivities. *J. Neurosci.* **18**, 10053–10069 (1998)
11. Stockman, A., Sharpe, L.T., Fach, C.C.: The spectral sensitivity of the human short-wavelength cones. *Vis. Res.* **39**, 2901–2927 (1999)
12. Stockman, A., Sharpe, L.T.: Cone spectral sensitivities and color matching. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 53–87. Cambridge University Press, Cambridge (1999)
13. Vos, J.J., Walraven, P.L.: On the derivation of the foveal receptor primaries. *Vis. Res.* **11**, 799–818 (1971)
14. Estévez, O.: Ph.D., Amsterdam University (1979)
15. Vos, J.J.: Colorimetric and photometric properties of a 2-deg fundamental observer. *Color. Res. Appl.* **3**, 125–128 (1978)
16. Stockman, A., MacLeod, D.I.A., Johnson, N.E.: Spectral sensitivities of the human cones. *J. Opt. Soc. Am. A* **10**, 2491–2521 (1993)
17. König, A., Dieterici, C.: Die Grundempfindungen und ihre Intensitäts-Vertheilung im Spectrum. In: *Sitzungsberichte Akademie der Wissenschaften in Berlin*, vol. 1886, pp. 805–829 (1886)
18. Brainard, D.H., Stockman, A.: Colorimetry. In: Bass, M., DeCusatis, C., Enoch, J., Lakshminarayanan, V.,

Li, G., Macdonald, C., Mahajan, V., van Stryland, E. (eds.) *The Optical Society of America Handbook of Optics. Vision and Vision Optics*, vol. III, 3rd edn, pp. 10.1–10.56. McGraw Hill, New York (2009)

Cone Fundamentals, Stockman-Sharpe

► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)

Cone Pigments

► [Photoreceptors, Color Vision](#)

Cone Spectral Sensitivities

► [Cone Fundamentals](#)

Congenital Color Vision Abnormality

► [Deuteranopia](#)
 ► [Protanopia](#)

Contextual Color Design

► [Environmental Color Design](#)

Continuous Ink Jet (CIJ)

► [Coloration, Ink-Jet Printing](#)

Contrary Colors

- [Complementary Colors](#)

Contrasting Colors

- [Complementary Colors](#)

Contribution of Melanopsin-Containing Retinal Ganglion Cell to Brightness Perception

- [Melanopsin Contributions to Human Brightness Perception](#)

Correlated Color Temperatures

- [Daylighting](#)

Cortical

- [Color Processing, Cortical](#)

Cortical Color Processing

- [Color Information Processing in Higher Brain Areas](#)

Crawford, Brian Hewson

Robert W. G. Hunt

Department of Colour Science, University of Leeds, Leeds, UK



1906–1991

Biography

Brian Hewson Crawford was a British physicist who made important contributions to the science of vision, colorimetry, lighting, and color rendering.

Brian Crawford was born in 1906 and died in 1991. His first scientific publication was, as his mother's ghostwriter, the Boy's Own Corner of the Daily Mail. He went to the University College London, graduating with first-class honors in Physics at the age of 19. After a brief spell at the Rodenside Laboratory of the photographic company, Ilford, he joined the staff of the National Physical Laboratory (NPL) in 1927.

Having very wide interests, Crawford explored languages, painting, and music. His attitude to life was well illustrated by his remark after breaking his wrist in a fall from a bicycle that it had a reset at a more convenient angle for playing the viola.

His honors included his Doctorate of Science from the University of London and his Newton Medal from the Colour Group (Great Britain).

Major Accomplishments/Contributions

Crawford worked in the NPL with W. S. Stiles under the leadership of John Walsh. A brilliant series of papers followed in the Proceedings of the Royal Society, on such new topics as equivalent backgrounds, increment thresholds, and, above all, the directional sensitivity of the retina to light and color known as the Stiles-Crawford effects. These were discovered during attempts to build a visual photometer based on pupillometry. The then “usual assumption that the apparent brightness of an object is proportional to the pupil area” was soon demolished [1–3].

Good fortune in research was characteristic of both Crawford and Stiles, but important discoveries fall only to those who deserve them. Crawford would say that you only had to search about in any field and you were bound to find out something interesting. No doubt this was so in his case. His 1947 *Proc. Roy. Soc.* paper on “Visual Adaptation in Relation to Brief Conditioning Stimuli” [4] is not obviously inspired by temporary blinding effects of gunfire flashes during the Second World War, but that was how “Crawford masking” was discovered. This is the effect whereby “a luminal stimulus begins to rise before the conditioning stimulus is applied to the eye.”

To compare and contrast the large-field colorimeters which Stiles and Crawford had each constructed in adjacent laboratories is an interesting exercise. Stiles’ machine was workshop built to the highest NPL precision and endowed with such facilities as a meteorology station of hygrometers, barometers, and thermometers, to monitor changes in the refractive index of the air. Crawford’s apparatus belonged to the string and sealing wax tradition, with corks for nonslip knobs and strips of graph paper for scales. Both instruments served their respective purposes admirably.

The successful early partnership and later divergence of Crawford and Stiles are perhaps explained by the creative tension between

opposite natures. To draw an analogy from art, Stiles’ science was classic; like Nicolas Poussin, he had neglected nothing. Crawford was a romantic; he was fond of quoting Maxwell’s dictum that it is always worth playing a trombone to a petunia at least once, and you never know what might happen. But Crawford could work to the highest NPL precision and accuracy when required, as he did in determining the average scotopic spectral response of the human eye, which forms the basis of the present CIE definition [5]. Not least was the difficulty of eliminating the effects of minute amounts of stray light and simultaneously discrediting data produced by several investigators who had not been so careful.

In his last years at NPL, Crawford studied the color rendering properties of artificial light sources, and he made a great breakthrough in convincing suspicious experts in art galleries and hospitals that certain fluorescent lamps were suitable for their exacting requirements [6]. Typical of Crawford was his finding that combinations of tungsten filament lamps and “radar blue” fluorescent lamps could imitate almost perfectly any phase of natural daylight. The resultant equipment was essential in providing a transportable reference illuminant for the darker corners of the Victoria and Albert Museum and the Sheffield Royal Infirmary. He also studied color matching and adaptation [7].

Work in the conservation department of the National Gallery was continued by Crawford long after officially retiring from NPL. He also made investigations on color in the laboratories of the University of Edinburgh, the Paint Research Association, Imperial College, and the Institute of Ophthalmology. He went on publishing original papers until the age of 79, thus refuting the fallacy that scientific research is only for young people [8].

References

1. Crawford, B.H.: The dependence of pupil size upon external light stimuli under static and variable conditions. *Proc. R. Soc. Med.* **B121**, 376 (1936)
2. Crawford, B.H.: The change of visual sensitivity with time. *Proc. R. Soc. Med.* **B123**, 69–89 (1937)

3. Crawford, B.H.: Photochemical laws and visual phenomena. *Proc. R. Soc. Med.* **B133**, 63 (1946)
4. Crawford, B.H.: Visual adaptation in relation to brief conditioning stimuli. *Proc. R. Soc. Med.* **B134**, 283 (1947)
5. Crawford, B.H.: The scotopic visibility function. *Proc. R. Soc. Med.* **B62**, 321 (1949)
6. Crawford, B.H.: Measurement of color rendering tolerances. *J. Opt. Soc. Am.* **49**, 1147 (1959)
7. Crawford, B.H.: Colour matching and adaptation. *Vision Res.* **5**, 71–78 (1965)
8. With the kind permission of the Colour Group (Great Britain), this account is based largely on the obituary written by Dr. D. A. Palmer and published in its newsletter. **17**, 1 (1992)

Cross-cultural Communication

Jules Davidoff
 Department of Psychology, Goldsmiths
 University of London, London, UK

Definition

Cultural variation and the role of the color lexicon in color tasks

Introduction

The pioneering studies of cross-cultural communication of color were those of Brown and Lenneberg [4], and its follow-up by Lantz and Steffle [11]. Their work came from the Whorfian perspective that language influences thought and was pursued, in the particular, by studying how color information can be communicated to another person. In their seminal investigation, Brown and Lenneberg [4] reported a positive correlation between a range of measures of codability for colors (speed of naming, consensus, communication accuracy) and the accuracy with which those colors were recognized. In particular, the focal or best example of a color is better remembered. However, Agrillo and Roberson [1] showed that the advantages found for coding and recognition of (focal) best examples of English

color categories result from the tight (Whorfian) links between linguistic and cognitive organization of categories, not from any inherent natural goodness of focal colors over other colors. When task demands are well-controlled, so that recognition need not be dependent on name retention, focal colors are no easier communicated than other colors. They argued that color terms define cognitive organization of colors; an idea that has proved useful even to the understanding the mechanisms whereby a population of autonomous agents (robots) could arrive at a repertoire of perceptually grounded categories that is sufficiently shared to allow successful communication [15].

Not all agree on the importance of color lexicons for communication about color. Lindsey, Brown, Brainard, and Apicella [12] observed a low consensus in color naming across Hadza observers, a remote hunter-gatherer society in Tanzania. They took this as evidence for a low evolutionary level of the Hadzane color lexicon. Most importantly, their study discovered a latent pattern of color naming across the speakers of the Hadzane language that did in fact reflect the English color lexicon, which is supposed to be a high evolutionary stage of color naming. Witzel [16] has suggested that the interpretation of the Hadzane data is complicated (confounded) by the variation in saturation of their color samples, but despite its importance, it is not pursued here. This chapter will not be much concerned with the origins of color categories but will consider the importance of color lexicons or more simply color names in color communication. It will show a complex picture. Gibson et al. [8] found that, when comparing the language of remote hunter-gatherers to Spanish, the variability of naming in the indigenous language leads to poorer communication for them about a colored stimulus though all languages communicated better about “warm” rather than “cool” colors. However, names were clearly of importance. Therefore, it is important to consider the precise task for which different cultures produce behavior that is or is not influenced by color lexicons. To start, the first question to consider is about behavior in

languages where colors are deemed to be similar enough to be given the same name.

Kay and Kempton (1984): Two Ways of Seeing Color Similarity

Brown and Lenneberg [4] found that the lack of terms in the Zuni language to distinguish between orange and yellow did not affect the ability to distinguish between colors in those regions. More recently, it has been shown (see [5]) that color categories do not derive from observers becoming more sensitive to colors at the boundary between two color terms. Thresholds and boundaries were comparatively assessed in English and Korean speakers. These languages differ in their color terms, but speakers all had the same threshold sensitivity irrespective of whether the color examined was at a boundary in one of the languages. These results bring us to the classic cross-cultural study of the different approaches to color similarity [9].

In Kay and Kempton, observers were presented with a triad of colors going from blue (A) to green (C) with an in-between color (B) that was half-way between the two in color space. However, (B) was chosen so that its color name was either that of (A) or (C) for English speakers. The importance of the study comes from the demonstration that observers showed two ways of making color judgments. If instructions were set to emphasize comparisons between two stimuli from the same color category, the observers did not find that (B) was more like (A) or (C). If the instructions were set so that the observer had to look at all three colors, then the odd-one-out chosen was the one from a different category – a result mostly replicated in other cross-lingual comparisons (see [5]). The observers were not told to name the colors to decide on their answers, though of course they could have. Kay and Kempton [9] report that observers did not name the colors, rather their choices were made, as one observer put it, according to the one “that looked different.” If so, the communication about the odd-one-out occurs because of a phenomenon known as categorical perception where color space has become

altered by the formation of a category, and so the color in the different category stands out. Most studies using the odd-one-out paradigm use only three colors, and so the categorical advantage could be explained from the observers naming the colors. More recently, studies have embedded the different color in many other identical colors, and hence the task becomes one of visual search. If latencies to perform such a task show a categorical advantage as well as short latencies, categorical perception can be inferred. No such studies have yet been published from remote cultures (see [5]).

The Kay and Kempton study also employed observers from a Mexican culture (Tarahumara) that has a *grue* language (same word for what in English is called green or blue) as well as English-speaking observers from the USA. It was only for the English speakers that there were two ways of responding. The critical proposal in Kay and Kempton [9] is that there are two ways of carrying out color decisions. Their argument is that both ways are perceptual, but one is based on categories and hence has potential for supporting the Whorfian position.

Recent physiological and neuroimaging research finds support for these two ways of making color similarity decisions [3]. Physiological evidence from the macaque for a red/green decision had suggested that the two ways are in competition with reciprocal activity of a population of cells encoding categorization and another population for discrimination. When the task changed, then the network of neuronal activity changed. A similar dual action has been found in human neuroimaging in a comparison between color naming and a task where attention was directed away from the same colors [3]. It was only when the task was color naming that a cortical area (in the temporal cortex around the human V4) was found that represented colors with greater within-category similarity. Subsequent studies (see [5]) employed a distracter task for which hue differences were irrelevant. Now, the brain area for which category difference was important, but hue distance unimportant, was found to be in the dorsolateral prefrontal cortices, whereas the visual cortex was the area interested in the metric

distance between colors. There was no category effect at any language related area, but one does have to ask how a frontal network for color categories was formed without language.

These recent findings relating brain area to the two ways of assessing color similarity raise important questions of the tasks used to investigate color similarity and the conditions that promote the use of a categorical code. The first of these tasks to be considered is simple and has a considerable history. It is the apparently uncontentious task of sorting colors.

Free-Sorting

Similarity and hence the number of categories could be determined by the closeness, in color space, of the to be sorted colors – this would be the physiological similarity way in Kay and Kempton. The physiological way could also be the categorical way if categories are determined by our physiology [10]. Boster [2] argued for this possibility in a task where English speakers were asked to sort eight colors (one example of black, white, red, yellow, green, blue, orange, and purple) in two groups and then to make further divisions to form three groups, four groups, and so on until there were eight groups each of one color. The performance of English speakers was found to mimic the evolutionary theory of Berlin and Kay. Boster's study has not been repeated for observers with smaller numbers of color terms.

Color sorting could, of course, also be carried out by Kay and Kempton's categorical way of seeing. Here one would normally assume that the color terms were learned and not innately (physiologically) driven. However, there are yet other explanations of color sorting that have come from cross-cultural studies. What to us are peculiarities in the free-sorting of color patches were noted by Luria [13] in his investigations of the Uzbeks, Luria was surprised at the allocation of their colors. Sorting did not depend upon what an English speaker would consider obvious. When asked why they had not put two orange-brown colors together, they said "It can't be done" "this is like calf's dung and this is like a peach; you

can't put them together." Here colors are considered not to be independent of objects an idea which has returned to those concerned with the origin of color categories [5]. What should also be noted is the wide variation in the number of color categories produced in the free-sorting task (see [5]). For example, a study of free-sorting 20 colors by speakers of Quechua (a five color-term language of Honduras) noted that they formed more groups than the terms in their language. However, studies of South African languages were less clear. A much more comprehensive study across a large number of languages was therefore undertaken by Roberson, Davies, Corbett, and Vandervyver [14]. Some important and perhaps surprising conclusions can be drawn from their data. They asked their observers to sort 65 stimuli drawn at roughly equal distances in color space between neighboring pairs of colors. If there is some underlying tendency in all observers to group certain colored stimuli together, then all observers should produce the same number of groups (categories) irrespective of the number of color terms in their language. What is immediately clear (see Table 1) is that this is not the case. Languages with around 11 of Berlin and Kay's basic terms (see [9]) do indeed produce a very similar number of groups; a result that might reflect the importance of those categories but could also be due to the similar color-cultural environments of first-world societies. However, the more interesting data in Roberson et al. [14] concern speakers of languages with few color terms; they produced much larger numbers of groups. Take, for example, the Himba – a people from remote Namibia – who have only 5 color terms but sorted the 65 stimuli into, on average, 21 groups. Do the Himba really have 21 color categories or are their responses being determined by other aspects of the task? Roberson et al. [14] noted many observers who formed groups based on brightness or saturation instead of hue; furthermore, sorting might be based on non-color factors. In the case of the Himba, recent research suggests that they are biased to concentrate on local aspects of a display [6]; this may explain a desire to accentuate dissimilarities between the colors and make a large number of small groups. Whatever

Cross-cultural Communication, Table 1 Basic colour terms are not important for communication from sorting. Roberson et al. [14]

Language	Sample size	Number of groups	Confidence limits $\pm 1.96SE$	Number of basic terms
Russian	75	13.5	1.72	12
Turkish	23	13.2	2.78	12
English 1	18	13.7	3.66	11
English 2	47	12.4	1.92	11
Tsakhur	19	11.3	2.16	11
Damara	40	10.5	0.43	11
Nama	56	15.7	1.82	11
Ndonga	80	13.5	1.10	11
Herero	20	16.4	2.94	11
Kwanyama	20	11.6	1.35	8
Bagwalal	25	33.6	4.07	7
Mbukushu	10	10.1	2.25	6
Kwangali	15	11.7	1.78	6
Tswana	44	17.3	1.92	6
Shona	17	12.4	1.61	6
Berinmo	17	20.3	0.69	5
Himba	21	21.4	4.10	5

the cause of their categorization performance, it is clear that neither the proposed universal color categories nor indeed their own color categories are a dominant force in their sorting behavior. Indeed, the same conclusion comes from the futile attempts of aphasic patients with normal color vision to sort colors (see [5]).

Matching-to-Sample

Another task where color categories (names) could inform a categorization task, but do not, is matching-to-sample. In a study comparing Himba with British observers [7], typical examples of two categories (e.g., green and blue) were presented, and the observer was asked which of these two was more similar to an intermediary color sample. The point of the study was to see if the sample was always matched to the color from the same category. Both for the Himba and the British sample, this was not the case. For example, British observers were shown typical examples of blue and green and asked to match a range of colors that were less typical examples of blue and green to one of the typical colors. The

range of intermediary colors was either centered at the blue/green boundary or shifted so that the middle of the range was clearly more like one of the two typical examples. A comparable task was given to the Himba using two of their color categories. Neither British nor Himba observers made use of color names/categories to make their decisions; both cultures matched equal number of samples to the two typical colors. Here a strategy of equal allocation overrode any decision based on color categories. Thus, Davidoff et al. [7] makes it clear that cultures may choose to ignore color categories even when they could be thought to be useful. It may even be the case that some cultures do not have the concept of color (see [5]). A famous comment of a Polynesian observer that “we don’t talk much about color here” makes the point that distinguishing between hues is not of interest to every culture.

Concluding Remarks

It is surprising that categories are not more important for communication in free-sorting and matching-to-sample if we consider our impression

of the color chart used by anthropologists/psychologists to study color terms. The immediate impression is not of colored stimuli seamlessly merging into those around them but of divisions within the display that conform to our color names. In the widely used display (see [5]), if asked exactly where the divisions occur, some deliberation might be required, but there is no doubt that the appearance is of a categorically organized display. So, it is reasonable to assume that color categories (and possibly their names) are of some importance to our processing and communication of colors. However, it is clear from Roberson et al. [14] and Davidoff et al. [7] that observers from many languages do not spontaneously name colors and/or use those names to communicate in color tasks. Of course, that only prompts the question of why some tasks/cultures are more inclined to use color names.

Cross-References

- [Ancient Color Terminology](#)
- [Bayesian Approaches to Color Category Learning](#)
- [Berlin and Kay Theory](#)
- [Bilingual Color Naming/Categories](#)
- [Color Categorical Perception](#)
- [Color Categories and Vagueness](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Color Category Learning in Naming-Game Simulations](#)
- [Color Dictionaries and Corpora](#)
- [Color Name Applications in Computer Vision](#)
- [Color Preference](#)
- [Cross-cultural Communication](#)
- [Crowdsourcing of Multilingual Color Names](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effects of Color Terms on Color Perception and Cognition](#)
- [Infant Color Categories](#)
- [Mechanisms of Color Categorization](#)
- [Memory Color](#)
- [Mesoamerican Color Survey Digital Archive](#)
- [Psychological Color Space and Color Terms](#)
- [Psychologically pure Colors](#)
- [Stroop Effect](#)
- [Translation of Color Language](#)
- [Unique Hues](#)
- [Vantage Theory of Color](#)
- [World Color Survey](#)

References

1. Agrillo, C., Roberson, D.R.: Colour language and colour cognition: Brown and Lenneberg revisited. *Vis. Cogn.* **17**, 412–430 (2009)
2. Boster, J.: Can individuals recapitulate the evolutionary development of color lexicons? *Ethnology*. **25**, 61–78 (1986)
3. Brouwer, G.J., Heeger, D.J.: Categorical clustering of the neural representation of color. *J. Neurosci.* **33**, 15454–15465 (2013)
4. Brown, R., Lenneberg, E.: A study in language and cognition. *J. Abnorm. Soc. Psychol.* **49**, 454–462 (1954)
5. Davidoff, J.: Color categorization across cultures. In: Elliot, A.J., Fairchild, M.D., Franklin A. (eds) *Handbook of Color Psychology*, pp. 259–278. Cambridge University Press: Cambridge (2015)
6. Davidoff, J., Fonteneau, E., Fagot, J.: Local and global processing: observations from a remote culture. *Cognition*. **108**, 702–709 (2008)
7. Davidoff, J., Goldstein, J., Tharp, I., Wakui, E., Fagot, J.: Perceptual and categorical judgements of colour similarity. *J. Cogn. Psychol.* **24**, 871–892 (2012)
8. Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S.T., Conway, B.R.: Color naming across languages reflects color use. *Proc. Natl. Acad. Sci.* **114**(40), 10785–10790 (2017)
9. Kay, P., Kempton, W.: What is the Sapir-Whorf hypothesis? *Am. Anthropol.* **86**, 65–78 (1984)
10. Kay, P., McDaniel, C.K.: The linguistic significance of the meanings of basic color terms. *Language*. **54**, 610–646 (1978)
11. Lantz, D., Steffire, V.: Language and cognition revisited. *J. Abnorm. Soc. Psychol.* **69**, 472–481 (1964)
12. Lindsey, D.T., Brown, A.M., Brainard, D.H., Apicella, C.L.: Hunter-gatherer color naming provides new insight into the evolution of color terms. *Curr. Biol.* **25**, 2441–2446 (2015)

13. Luria, A.R.: Cognitive Development: Its Cultural and Social Foundations. Harvard University Press, Cambridge (1976)
14. Roberson, D., Davies, I.R.L., Corbett, G.G., Vandervyver, M.: Free-sorting of colors across cultures: are there universal grounds for grouping? *J. Cogn. Cult.* **5**, 349–386 (2005)
15. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language: a case study for color. *Behav. Brain Sci.* **28**, 469–485 (2005)
16. Witzel, C.: New insights into the evolution of color terms or an effect of saturation? *Iperception*. **7**, 2041669516662040 (2016)

Crowdsourcing of Multilingual Color Names

Dimitris Mylonas

Department of Computer Science, University College London, London, UK

Synonyms

[Collective intelligence](#); [Color language](#); [Online experiments](#)

Definition

Crowdsourcing of multilingual color names refers to distributed experimental methodologies designed to collect large datasets of color naming responses in different languages from a large number of observers over the Internet.

Color Naming

Color naming links vision and language. It describes the intriguing cognitive capacity of humans to communicate about regions of color space using category labels, for example, yellow, navy blue, and dark olive green [1]. A small number of basic color terms, corresponding to the English white, black, red, yellow, green,

blue, brown, orange, pink, purple, and grey, are shared and comprehended well by most speakers in many languages [2]. Yet, color names vary across languages, lexically, in number and in range of reference. To augment color communication within and between different languages, it is necessary to have a worldwide method for gathering the words and phrases that people use to describe a wide range of colors [3, 4].

Color systems are usually three-dimensional geometric spaces that allow the description of colors with numerical coordinates. The representation of perceived colors in such spaces can be very useful for specifying the referents of color names in terms of their coordinates. Using color systems in cross-cultural research enables measurements of differences and similarities of color categories across languages. In Fig. 1 are shown subdivisions of a continuous color plane into English and German lexical color categories, based on the responses of human observers in a color naming experiment [5]. It is interesting that not only the names are different (language-dependent) but also the boundaries of the corresponding regions. For example, the region for *hellblau* in German is larger than the region for *light blue* in English.

Crowdsourcing and Laboratory Color Naming Studies

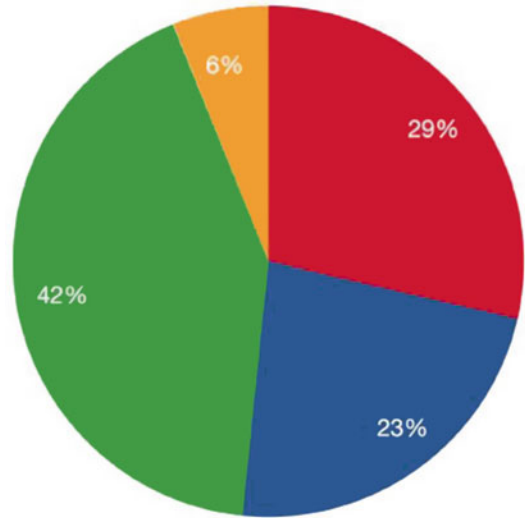
Human-centered color naming studies are traditionally performed in controlled laboratory settings and involve predefined naming lists (e.g., restricting the responses to a limited number of single word terms) and labor-intensive tasks for a small number of observers [6–8]. Despite the usefulness of calibrated experiments in the field of color science, experimenters face several limitations. First, the controlled viewing conditions, while advantageous for accurate colorimetric specification of stimuli, limit the validity of the predicted color naming functions in real-world settings. Second, the pool of available observers is often constrained to a small number of college



Crowdsourcing of Multilingual Color Names, Fig. 1 Classification synthetic image (left) by observer-based color names in English (middle) and German (right). Each category is represented by the color of its centroid [5]

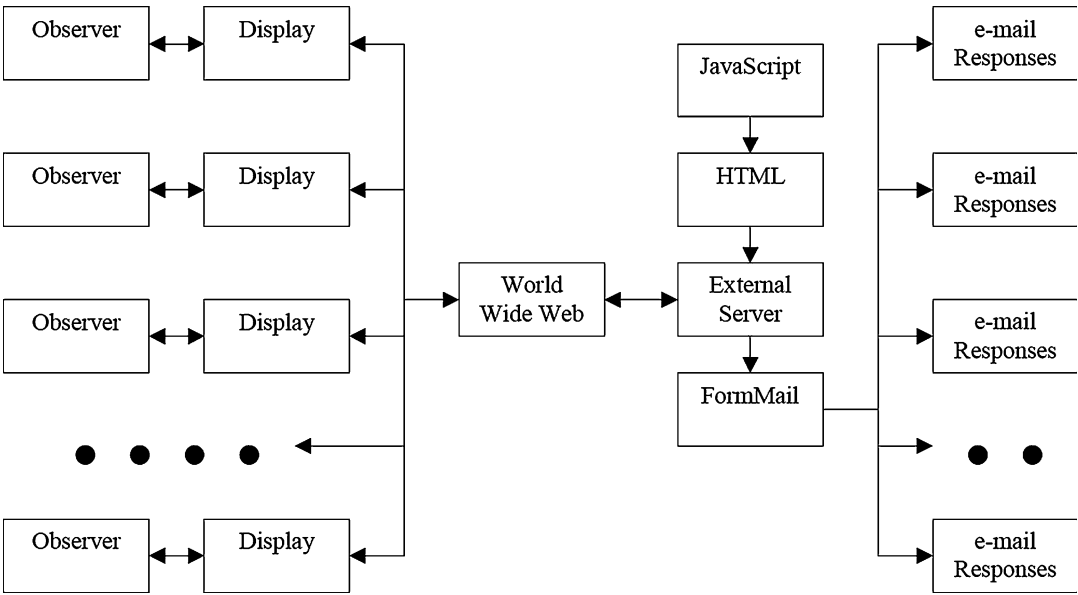
students or to the authors of the study, which makes it difficult to generalize the results to heterogeneous groups of the population. Third, the large individual differences reported in color naming studies makes the generalization of laboratory experimental findings open to doubt. Also, constrained color naming methods are able to capture only a small fraction of the richness of color languages of the world and the labeling of large regions of color space remains ambiguous. For example, Fig. 2 shows the percentage of number of words used by subjects in an unconstrained color naming experiment [4]. Constraining the responses only to the 11 basic color terms would account for 29% of the responses while to single words would account only for 52% of the responses for naming stimuli across the full color gamut.

On the other hand, web-based color naming experiments provide greater ecological validity than traditional approaches by allowing simultaneous participation of observers in their own familiar spaces, in their own time, with their own equipment and without the physical attendance of the examiner [4, 9]. They also avoid the restrictions of the usual methods which have a small number of observers and/or a limited set of monolexic color terms. Instead, thousands of observers from linguistically and demographically diverse populations are able to name freely a large number of colors online and produce larger color lexicons that improve the precision of color names. Online methods also depart from previous



Crowdsourcing of Multilingual Color Names, Fig. 2 Number of words used by subjects in an unconstrained color naming experiment in English. In red proportion of 11 basic color terms, in blue single words non-basic terms, in green two-word color names and in orange color names consisting of three-words or more [4]

research by distributing the color naming task over a large population. In the collection of multilingual color naming data, online experiments extend earlier cross-cultural studies, which typically used only the most saturated color samples on the surface of the Munsell system, by also sampling the interior of the color solid [2, 3]. Yet, online experimental methodologies often receive criticism as not meeting the exacting standards demanded for rigorous color research



Crowdsourcing of Multilingual Color Names, Fig. 3 Overall flowchart of web-based color naming experiment [9]

because of the uncalibrated color reproduction and viewing conditions. For example, different participants may use different display technologies and browsers and they may view the color stimuli from different distances and viewing angles under various illumination conditions [9]. Furthermore, the motivation of participants varies in both reward-based and in volunteering-based crowdsourcing, and screening the quality of the responses is a major challenge. However, the practical success of sRGB in color encoding for Internet applications has greatly improved the consistency of displayed colors and people exhibit a remarkable ability to describe consistently the color of objects in a continually changing visual world. The robustness of crowdsourcing experimental results can be assessed by whether or not their variability is larger than the variability in laboratory-based studies [4, 7, 9].

Crowdsourcing Color Naming Experiments

To understand the workings of crowdsourcing, a number of online color naming experiments are

briefly reviewed. Nathan Moroney [9] was the first to publish the results of an unconstrained web-based color naming experiment. More than 700 participants were asked to give the best names for seven patches of color selected randomly from a $6 \times 6 \times 6$ uniform grid sampling of the RGB cube, viewed on their own displays against a white background. Figure 3 shows a simplified flowchart of this experiment to create a dynamic website that presents colors and collects color naming responses.

Findings of the online experiment of Moroney [9] were validated against the results of Boynton and Olson [6] and Sturges and Whitfield [8], both obtained under controlled laboratory conditions. The comparisons showed a high degree of correlation for the chromatic basic color terms in terms of hue angles in CIELAB with a linear fit of $R^2 = 0.99$ and $R^2 = 0.98$, respectively. The same comparison but for 27 common color names against the results of a subsequent web-based study resulted in a linear fit of $R^2 = 0.99$ [4]. In 2008, DoloresLabs publicized the results of an online color naming survey conducted via Mechanical Turk to collect a small number of responses for a large number of colors. In the

raw dataset, a total of 1966 distinct color descriptions were collected for 10,000 randomly sampled colors in the RGB cube. An analysis of the DoloresLabs dataset by [10] showed a good correspondence for colors on the surface of the Munsell system with the results of the World Color Survey (WCS) for red, brown, and purple but hue shifts for blue and green [3]. The analysis of data across the full color gamut revealed two additional consistent categories at the lavender and cyan regions that do not correspond to any salient region in the WCS data (Fig. 4).

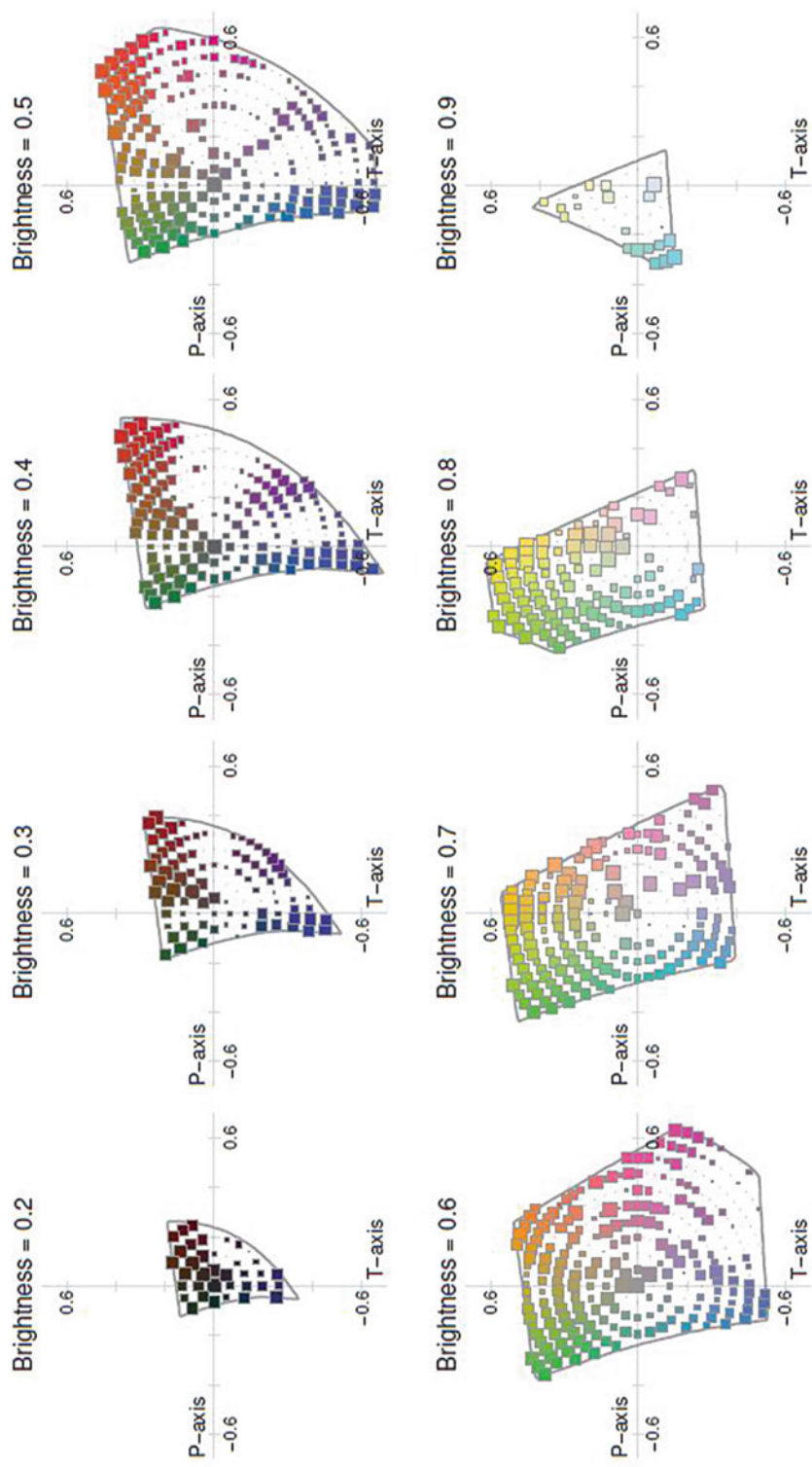
In 2010, the author of the web comic XKCD, Randall Munroe collected a dataset of 3.4 million unconstrained responses mainly in English. This very large dataset is associated with metadata regarding the gender, language skills, color-blindness of the subjects, and information about their monitor settings. Observers were free to name as many sets of color swatches as they liked, presented against a white background. Each color swatch was uniformly sampled from the full RGB cube. An analysis of the publicly available dataset by Heer and Stone [11] showed clusters that correspond well to the basic color terms identified by Berlin and Kay [2].

In 2009, Mylonas and MacDonald launched an ongoing color naming experiment to collect broad sets of color names within different languages with their corresponding color ranges in sRGB and Munsell specifications over the Internet [4]. Each observer was presented with a sequence of 20 color patches randomly selected from 600 simulated Munsell samples, presented against a neutral grey background. The color naming responses were associated with metadata regarding the response time, color deficiency, hardware/software components, viewing conditions, gender, and cultural background of the observers. An analysis of the data for the most important color names suggested the extension of the English inventory from the 11 basic color terms of Berlin and Kay [2] to 13 basic terms including a lilac and a turquoise category ([12]; see also [10]). Except blue, comparisons of

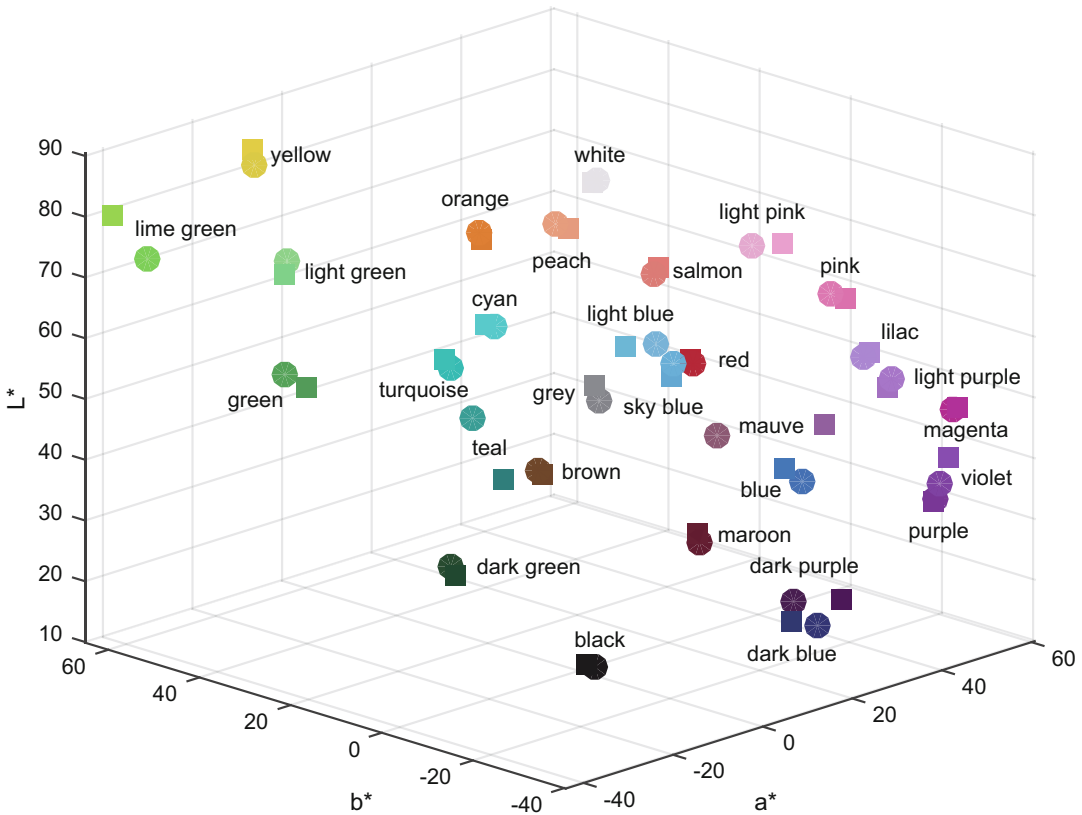
centroid location of the basic color terms in different languages showed a good correspondence ($>5\Delta E_{00}$) despite the online experimental methodology and the linguistic diversity of the observers [5, 13]. However, the agreement between British and American English ($\Delta E_{00} = 1$) was better than the agreement between English and non-English color languages [14]. Figure 5 shows the location of the thirty most frequent color names offered by a large number of British and American English speakers in the online color naming experiment. The online experiment was also validated directly against an unconstrained color naming experiment using the same stimuli on a calibrated CRT monitor in controlled viewing conditions [7]. The agreement ($<5\Delta E_{00}$) for the location of the basic color terms between the web- and laboratory-based experiments was satisfactory and superior to the agreement ($>7\Delta E_{00}$) between previous laboratory-based studies [6, 8].

Conclusions

Crowdsourcing experimental methodologies offer considerable advantages over traditional approaches for obtaining richer color naming datasets in multiple languages. Comparisons between the results of web- and laboratory-based studies show that the locations of color categories correspond well and support the validity of both methods in estimating color naming functions in calibrated and uncalibrated monitor settings. Moreover, the overall good agreement between the location of corresponding basic color terms in multiple languages confirms that speakers of different languages tend to categorize colors into basic color terms similarly but differences do exist, especially in the blue region and between nonbasic categories. New mobile display technologies and online platforms offer new capabilities and present new challenges to researchers for conducting color naming research over the Internet.



Crowdsourcing of Multilingual Color Names, Fig. 4 Consistency of color naming responses on 8 brightness levels in IPT space. The size of each colored square corresponds to the degree of agreement between observers for the name of this chip [10]



Crowdsourcing of Multilingual Color Names, Fig. 5 Location of centroids of 30 most frequent color names in CIELAB for British (square) and American (circle) English speakers [14]

Cross-References

- [Cross-cultural Communication](#)
- [Psychological Color Space and Color Terms](#)

References

1. MacDonald, L., Mylonas, D.: Colour Naming – Linking Vision and Speech, pp. 16–25. Krakow (2018)
2. Berlin, B., Kay, P.: Basic Color Terms: their Universality and Evolution. Center for the Study of Language and Information, Stanford (1969/1991)
3. Cook, R.S., Kay, P., Regier, T.: The world color survey database. In: Cohen, H., Lefebvre, C. (eds.) Handbook of Categorization in Cognitive Science, pp. 223–241. Elsevier, Amsterdam (2005)
4. Mylonas, D., MacDonald, L.: Online colour naming experiment using Munsell samples. In: 5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science. CGIV 2010/MCS’10, Joensuu (2010)
5. Mylonas, D., Stutters, J., Doval, V., MacDonald, L.C.: A synthetic observer for colour communication. In: MacDonald, L., Westland, S., Wuerger, S. (eds.) Proceedings of 12th Congress of the International Colour Association, pp. 701–704 (2013)
6. Boynton, R.M., Olson, C.X.: Locating basic colors in the OSA space. *Color Res. Appl.* **12**(2), 94–105 (1987)
7. Mylonas, D., Griffin, D.L., Stockman, A.: Mapping colour names in cone excitation space. In: 25th Symposium of the International Color Vision Society, Riga (2019)
8. Sturges, J., Whitfield, A.: Locating basic colours in the Munsell space. *Color. Res. Appl.* **20**(6), 364–376 (1995)
9. Moroney, N.: Unconstrained web-based color naming experiment. In: Eschbach, R., Marcu, G.G. (eds.) Color Imaging VIII: Processing, Hardcopy, and Applications. Proceedings of the SPIE, vol. 5008, pp. 36–46 (2003)
10. Chuang, J., Hanrahan, P., Stone, M.: A Probabilistic Model of the Categorical Association Between Colors,

- pp. 6–11. The Society for Imaging Science and Technology, Portland (2008)
11. Heer, J., Stone, M.: Color naming models for color selection, image editing and palette design. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 1007–1016 (2012)
 12. Mylonas, D., MacDonald, L.: Augmenting basic colour terms in English. *Color Res. Appl.* **41**(1), 32–42 (2016)
 13. Paramei, G.V., Griber, Y.A., Mylonas, D.: An online color naming experiment in Russian using Munsell color samples. *Color Res. Appl.* **43**(3), 358–374 (2018)
 14. Mylonas, D., MacDonald, L., Griffin, L.D.: Differences in color naming between British and American English speakers. In: Proceedings of 13th AIC Congress. Jeju (2017)

Culturally Specific Basic Color Terms

► [Color Categorization: Patterns and Mechanisms of Evolution](#)

Cyan, Magenta, Yellow, or Black (CMYK)

► [Coloration, Ink-Jet Printing](#)

D

Dalton, John

Stephen Westland
Colour Science and Technology, University of
Leeds, Leeds, UK



Biography and Major Accomplishments/Contributions

John Dalton was an English chemist, physicist, and meteorologist who is best known for his work in the development of modern atomic theory and his research into color blindness.

Dalton was born in the county of Cumberland in England in 1766 and received his early education from his father who ran a private Quaker school. He started his own school aged 11, and at age 15 he joined his brother teaching in a school in Kendal [1]. He was barred from attending English universities

because he was a dissenter (a Christian opposed to state religion and mandatory membership in the Church of England) but acquired scientific knowledge from John Gough, a blind philosopher with wide-ranging scientific interests. However, in 1793, he was appointed as teacher of mathematics at New College, Manchester (a dissenters' college). On 3 October 1794, Dalton was elected as a member of the Manchester Literary and Philosophical Society, and a month later he presented his first paper entitled, *Extraordinary Facts Relating to the Vision of Colors with Observations*. The society's *Memoirs and Proceedings* has been published continuously since its first edition which, when it was launched in 1783, was the only regular scientific journal in the United Kingdom except for the *Philosophical Transactions of the Royal Society*. As a result of this paper, congenital color vision deficiency is still referred to as "Daltonism" because he was the first to describe the condition in detail, and the paper also stimulated great debate among other investigators [1]. In 1800 he left New College and began private teaching, and in 1808 he published the atomic theory for which he is most famous. He became President of the Manchester Literary and Philosophical Society in 1817, a post he retained until his death in 1844.

Dalton's Color Blindness

Dalton noticed that only males were afflicted with his condition, and we now know that color

blindness, or color deficiency as it may more properly be referred to, is much more prevalent among males than females. He also recognized that the condition must be hereditary since he and his brother had the same condition.

In his 1794 letter, Dalton wrote:

The flower was pink but it appeared to me almost an exact sky-blue by day; in candlelight however it was astonishingly changed, not having then any blue in it but being what I call red – a colour which forms a striking contrast to blue.

Thirty-six years prior to Dalton's famous 1794 letter, Thomas Young had postulated that congenital color vision defects arose from the photoreceptors, but Dalton refused to believe it [1]. As a result of his meticulous observations, Dalton believed that the fluids in his eyes must contain a blue colorant, and he instructed that on his death his eyes be subject to a postmortem analysis. No trace of the blue colorant was found. Interestingly Dalton found 25 others (including, notably, his own brother) who suffered the same failure of color constancy. Dalton also wrote that the red end of the spectrum was "little more than a shade or defect of light."

We now know that Young was essentially correct, that is, that congenital color vision defects result from genetic mutations that mean that one (or more) of the three visual pigments found in the cones of our retina has anomalous spectral absorption properties or is absent entirely. Young believed that Dalton was a protanope – that is, that he was missing the long-wavelength sensitive visual pigment. Although the postmortem examination of Dalton's eye did not reveal the blue colorant that Dalton had predicted, the remaining eye was preserved between two sheets of glass until it was analyzed by Mollen et al. who concluded that Dalton was indeed a dichromat but was a deuteranope rather than a protanope [2].

Dalton's theory that color deficiency was caused by a long-wavelength absorbing colorant in the fluids of his eyes ultimately proved to be incorrect. He was also not the first to "discover" color blindness. However, he made a substantial contribution to the understanding of color blindness through his meticulous observations, and he also inspired other scientists to study this fascinating phenomenon.

Atomic Theory

Although Dalton's name is indelibly associated with color deficiency, by far his most important work was in the area of atomic theory which he formally first published in 1805. The main points of Dalton's atomic theory were:

1. Elements are made of extremely small particles called atoms.
2. Atoms of a given element are identical in size, mass, and other properties; atoms of different elements differ in size, mass, and other properties.
3. Atoms cannot be subdivided, created, or destroyed.
4. Atoms of different elements combine in simple whole-number ratios to form chemical compounds.
5. In chemical reactions, atoms are combined, separated, or rearranged.

There is uncertainty about how Dalton arrived at his atomic theory [3]. Also, there was no evidence available at the time to scientists to deduce how many atoms of each element combine to form compound molecules so that Dalton, for example, wrongly assumed that water was composed of a single oxygen atom combined with a single hydrogen atom. Nevertheless, the essential features of his theory have survived to this day, and his theory can be said to be one of the bedrocks of modern chemistry.

Other Contributions

Dalton made notable contributions to meteorology (where he made over 200,000 observations, published several important essays on the gas laws (his law of partial pressures became known as Dalton's law), and wrote about rain, the color of the sky, and refraction).

References

1. Dickinson, C., Murray, I., Carden, D.: Dalton – letter. In: Dickinson, C., Murray, I., Carden, D. (eds.) *John Dalton's Colour Vision Legacy*. Taylor & Francis, London (1997)

2. Mollon, J.D., Dulai, K.S., Hunt, D.M.: Dalton’s colour blindness: an essay in molecular biography. In: Mollon, J.D., Dulai, K.S., Hunt, D.M., Dickinson, C., Murray, I., Carden, D. (eds.) *John Dalton’s Colour Vision Legacy*. Taylor & Francis, London (1997)

3. Nash, L.K.: The origin of Dalton’s chemical atomic theory. *Isis*. **47**(2), 101–116 (1956) The History of Science Society CrossRef

indirectly from a sky of assumed or known luminance distribution to the illuminance on a horizontal plane due to an unobstructed hemisphere of this sky, where the contribution of direct sunlight to both illuminances is excluded [1].

Average Daylight Factor

The daylight factor provides the potential of the available daylight in interiors. It is defined as the ratio of the illuminance at a point in an interior room on a given plane due to the light received from a sky of a defined luminance distribution, excluding direct sunlight ($E_{\text{plane, room, CIE overcast}}$), to the illuminance on a horizontal plane in the open outside, without obstructions ($E_{\text{plane, outside, CIE overcast}}$). Usually, the CIE overcast sky and the horizontal plane are used as the standard situation. The daylight factor is expressed in percentages:

$$DF = 100 \cdot \frac{E_{\text{hor, room, CIE overcast}}}{E_{\text{hor, outside, CIE overcast}}}$$

Both the average daylight factor and the daylight factor distribution in the space are of interest. The average daylight factor is often a good predictor of the overall appearance of a room. Tregenza and Wilson [2] provided for offices a qualitative description of room appearance for different values of the average daylight factor (Table 1).

Daltonism

- Deuteranopia
- Protanopia

Daylight

- Daylighting

Daylight Factor

Wout van Bommel
van Bommel Lighting Consultant, Nuenen, The Netherlands

Definition

Ratio of the illuminance at a point on a given plane due to the light received directly and

Daylight Factor, Table 1 Average daylight factor in offices with side windows and the corresponding visual character of the room

Average daylight factor	Room character	Effect of electric lighting
1%	Gloomy, harsh contrasts when viewing out	Electric lighting may mask daylight variation
2%	Areas deeper in the room may seem underlit	Appearance of daylit room even if electric lighting is the main illumination
5%	Looks brightly daylit; visual and thermal discomfort may occur with large window areas	Electric lighting rarely needed
10%	Looks like a semi-outdoor space; visual and thermal conditions may be unsuitable for office-type tasks	

Source: [2]

Average daylight factors between 2% and 5% usually correspond to acceptable to good visual conditions.

Graphical methods have been developed in the before-PC era to determine the daylight factor for buildings depending on their layout and fenestration. Today, most lighting calculation software packages, often available for free, can calculate daylight factors. A relatively good and useful approximation of the average daylight factor, at least for the critical range of daylight factors between 1 and 5, is obtained from the ratio of the sum of window areas and the sum of the surface areas of the room (floor, ceiling, and wall surfaces). Figure 1 (scale on the left) gives the correlation between daylight factors determined for many different situations and this area ratio [2].

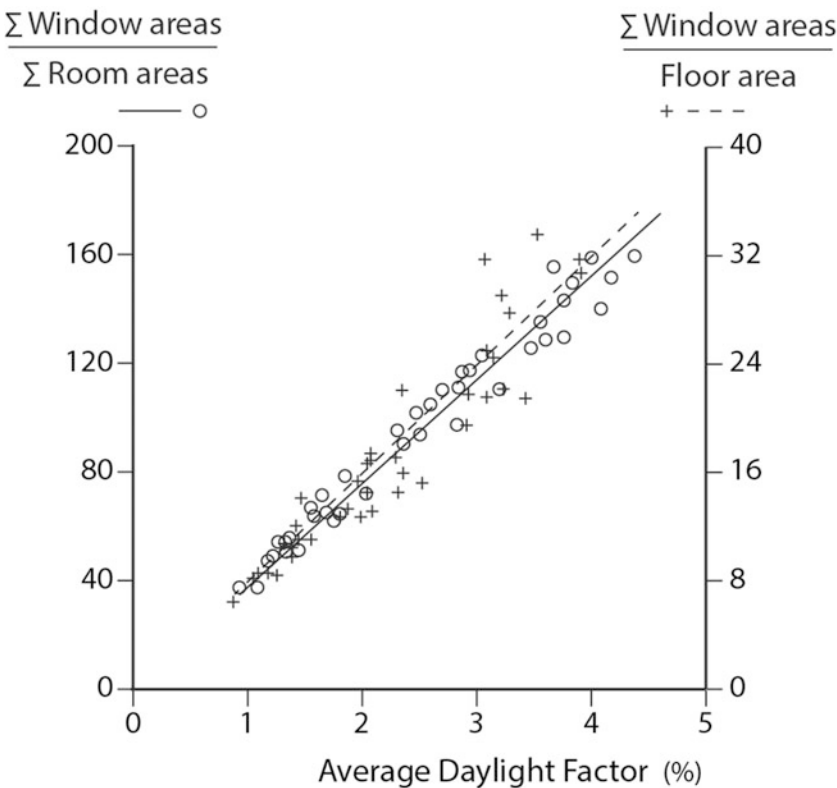
It shows that the daylight factor can be predicted with high accuracy (for the range between 1% and 5%) with:

$$DF_{av} \approx \frac{40 \times \sum \text{Window areas}}{\sum \text{Room surface areas}}$$

Another, even more convenient predictor is the ratio of the sum of window areas and the floor area. The correlation, only slightly less accurate, is given in Fig. 1 with the scale on the right [2]. The predictor (again for the range between 0 and 5) is:

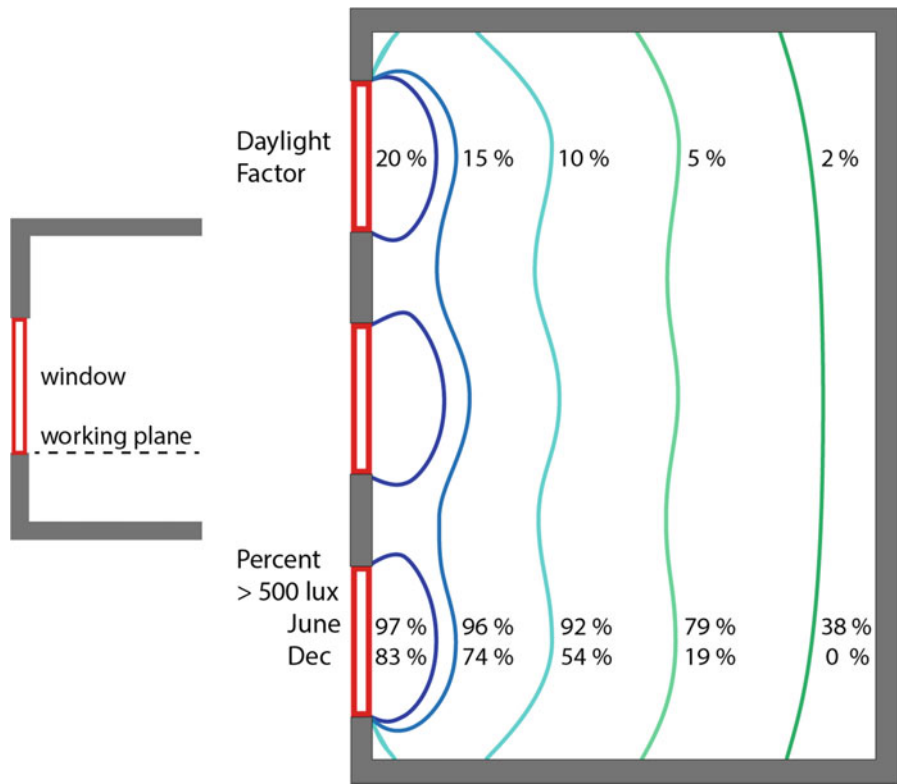
$$DF_{av} \approx \frac{8 \times \sum \text{Window areas}}{\text{Floor area}}$$

Some building standards use the ratio of window areas and floor area as a criterion to safeguard sufficient daylight entrance in buildings.



Daylight Factor, Fig. 1 Correlation between the average daylight factor of a room and the ratio of the sum of the window areas and the sum of the room surface areas (scale

on the left) and the ratio of the sum of the window areas and the floor area (scale on the right). (Source: [2])



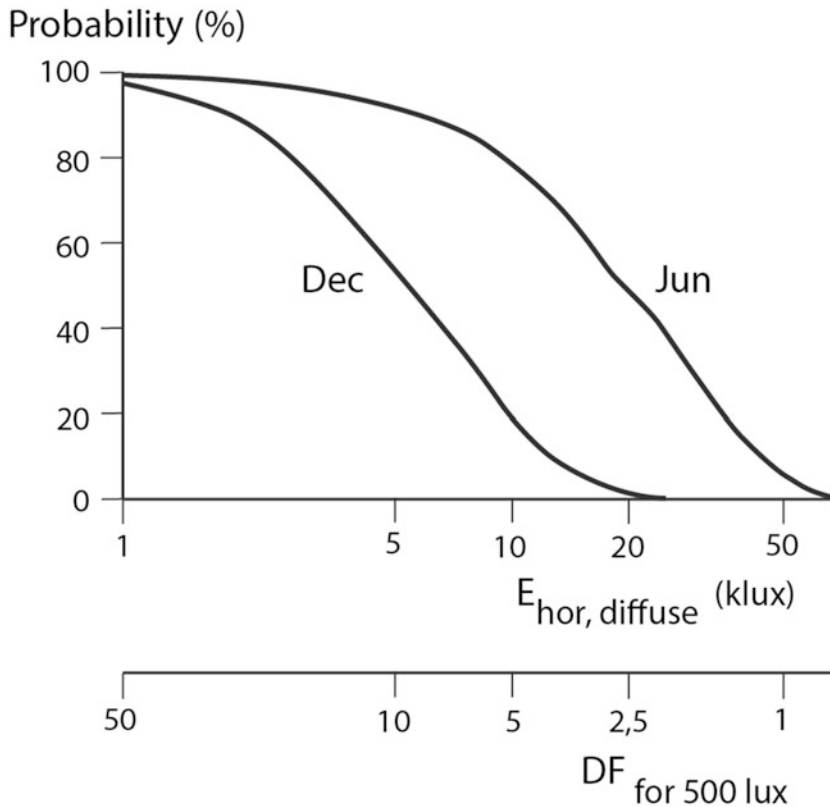
Daylight Factor, Fig. 2 Daylight factor contours for an office of 5.4×3.6 m with window height and vertical position as shown in the figure and for surface reflectances for ceiling, walls, and floor of, respectively, 0.7, 0.5, and

0.2. At the bottom, the corresponding percentages of the time that more than 500 lux will be obtained are given for June and December (based on daylight measurements in Bratislava)

Daylight Factor at a Point

The daylight factor at a point in space gives information about the potential of daylight at different locations in a room. As an example, Fig. 2 gives an example of daylight factor contours for a typical one- or two-person office room. The daylight factor at a specific point in the room gives the daylighting potential of a building at that point for the poorest type of sky represented by the CIE standard overcast sky. What the actual illuminance at that point will be depends on the time of the day, the time of the year, and the actual location on earth (sun position). For some locations on earth, historical data is available on the probability that a given horizontal illuminance as a result of

skylight (excluding direct sunlight) is exceeded [3–5]. Figure 3 shows, as an example, a graph with such probability for Bratislava (geographical latitude approx. 48°N). Using the example of Bratislava, the daylight factors of the contours of Fig. 2 have been converted into percentages of the time that more than 500 lux horizontal illuminance is obtained in June and December, respectively, in Bratislava (values at the lower end of Fig. 2). Where in the summer month daylight can provide a substantial part of the required 500 lux, this is in the winter month impossible, even close to the windows. Of course, because of the sun’s position, at higher latitudes, lower lighting levels will prevail and at lower latitudes higher levels.



Daylight Factor, Fig. 3 The probability that a given horizontal illuminance due to skylight is exceeded in Bratislava in December and June, respectively; source, [4];

second horizontal scale, daylight factor (DF) required for an interior lighting level of 500 lux horizontally

Cross-References

► [Daylighting](#)

References

1. CIE Publication: CIE DIS 017/E:2016 ILV: International Lighting Vocabulary, 2nd edn. Vienna (2016)
2. Tregenza, P., Wilson, M.: Daylighting, Architecture and Lighting Design. Routledge, Abingdon (2011)
3. Muneer, T.: Solar Radiation and Daylight Models, 2nd edn. Elsevier-Butterworth-Heinemann, Oxford (2004)
4. Kittler, R., Kocifaj, M., Darula, S.: Daylight Science and Daylighting Technology. Springer, New York/Dordrecht/Heidelberg/London (2012)
5. Mardaljevic, J., Christoffersen, J.: 'Climate connectivity' in the daylight factor basis of building standards. Building and Environment. **113**, 200–209 (2017)

Daylight Illuminants

Balázs Kránicz

University of Pannonia, Veszprém, Hungary

Definitions

Illuminants having the same or nearly the same relative spectral power distributions as phases of daylight.

Description of Daylight Phases

Originally, CIE standardized three illuminants for colorimetric purposes, illuminants A, B, and C. Illuminant C was the illuminant to represent

daylight. CIE also defined standard sources for these illuminants, and CIE standard source C was realized by using an incandescent lamp and some liquid filters. This source had, however, less radiation in the ultraviolet spectrum as natural daylight. With the increased use of optical brighteners, it became necessary to define daylight that came nearer to natural phases of daylight.

In the 1960s, the question how daylight power distributions of different correlated color temperatures could be calculated by a simple and explicit formula came to the front. The task was to determine relative spectral power distributions of daylight that could be measured on the surface of the Earth and was modified by the scattering and filtering effect of the sky.

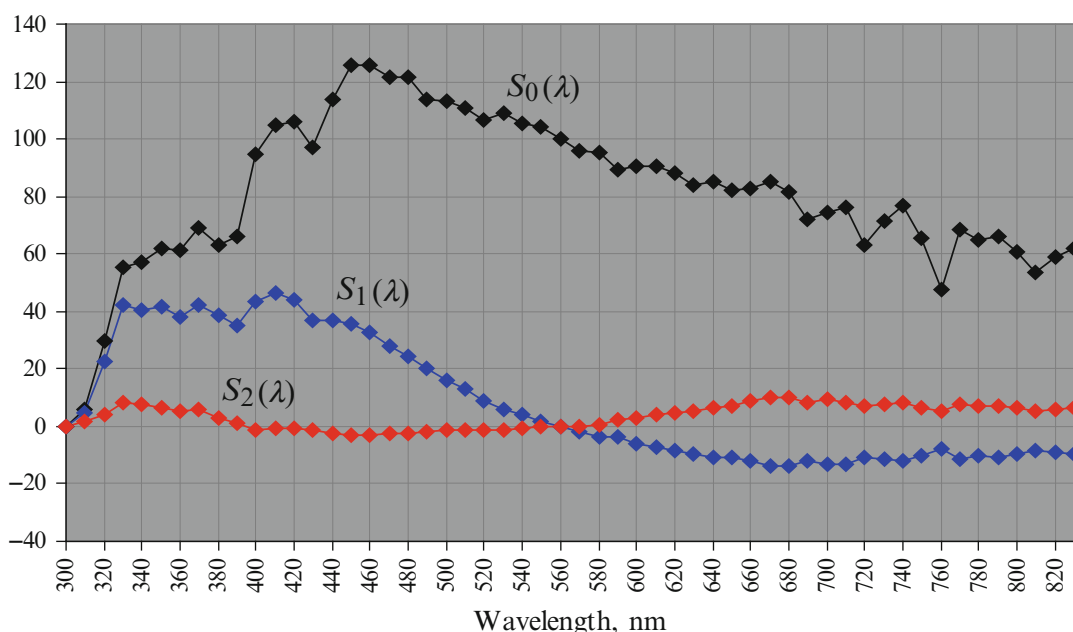
Pioneer importance can be assigned to the publication of Judd, Wyszecki, and MacAdam [1]. The authors and their coworkers performed spectroradiometric measurements, and with proper extrapolation of the wavelength set of a number of observations, a data base was completed from 300 to 830 nm with 10 nm steps.

From the measurement data, the average spectral power distribution (in references mostly denoted by $S_0(\lambda)$) and two characteristic distributions ($S_1(\lambda)$ and $S_2(\lambda)$) were determined; see Fig. 1. The authors themselves called the method applied *characteristic vector method*, and it can be considered as a kind of *principal component analysis*. Of course, more than two characteristic vectors had been determined, but it was obvious that for a given chromaticity coordinate pair, the two abovementioned characteristic distributions could be used. (In principal component analysis, they belong to the two biggest eigenvalues of the covariance matrix of the data.)

Based on the mean vector $S_0(\lambda)$ and the two most important characteristic vectors or distributions, daylight power distribution $S(\lambda)$ of correlated color temperature T_{cp} can be calculated as follows:

$$S(\lambda) = S_0(\lambda) + M_1 \cdot S_1(\lambda) + M_2 \cdot S_2(\lambda). \quad (1)$$

Factors M_1 and M_2 in the formula above are not to be handled as constants; hence, these values



Daylight Illuminants, Fig. 1 The mean vector and the first two characteristic vectors to reconstitute spectral power distributions of daylight

depend on chromaticity coordinates x_D , y_D belonging to the given correlated color temperature T_{cp} :

$$M_1 = \frac{-1,3515 - 1,7703 \cdot x_D + 5,9114 \cdot y_D}{0,0241 + 0,2562 \cdot x_D - 0,7341 \cdot y_D} \quad (2)$$

and

$$M_2 = \frac{0,0300 - 31,4424 \cdot x_D + 30,0717 \cdot y_D}{0,0241 + 0,2562 \cdot x_D - 0,7341 \cdot y_D} \quad (3)$$

where in the case of $4,000 \leq T_{cp} < 7,000$

$$x_D = -4,6070 \cdot \frac{10^9}{T_{cp}^3} + 2,9678 \cdot \frac{10^6}{T_{cp}^2} + 0,09911 \cdot \frac{10^3}{T_{cp}} + 0,244063 \quad (4)$$

and in the case of $7,000 \leq T_{cp} \leq 25,000$

$$x_D = -2,0064 \cdot \frac{10^9}{T_{cp}^3} + 1,9018 \cdot \frac{10^6}{T_{cp}^2} + 0,24748 \cdot \frac{10^3}{T_{cp}} + 0,237040. \quad (5)$$

In both correlated color temperature ranges

$$y_D = -3,000 \cdot x_D^2 + 2,870 \cdot x_D - 0,275. \quad (6)$$

The unit of quantity of correlated color temperature T_{cp} in conditions of formula Eqs. 4 and 5 is kelvin (K). The form of Eqs. 4, 5, and 6 and the parameters in them come from curve fitting. The purpose of the curve fitting was to approximate the functional relation between correlated color temperature values and chromaticity coordinates of spectral power distributions of the measurement data set.

On the basis of the facts above, chromaticity coordinates, x_S and y_S , of $S(\lambda)$ determined by formula Eq. 1 have to match chromaticity coordinates x_D and y_D (from Eqs. 4, 5, and 6) belonging to correlated color temperature T_{cp} .

Nominal and Actual Temperature

The CIE recommended daylight illuminants in 1967. The correlated color temperatures are affected by the numerical value of the radiation constant c_2 . In accordance with the International Practical Temperature Scale, 1948, amended in 1960, which was in use at the time when the procedure for calculating daylight illuminants was adopted by the CIE, the value of c_2 was equal to $1,438\,0 \times 10^{-2} \text{ m} \cdot \text{K}$. With this value, the correlated color temperature of illuminant D65 was exactly equal to 6,500 K. The change of c_2 to the value of $1,438\,8 \times 10^{-2} \text{ m} \cdot \text{K}$ (International Practical Temperature Scale, 1968) increases the correlated color temperatures of illuminant D65 by the factor $1,438\,8/1,438\,0$. Thus, the correlated color temperature increases by approximately 4 K [2, 3].

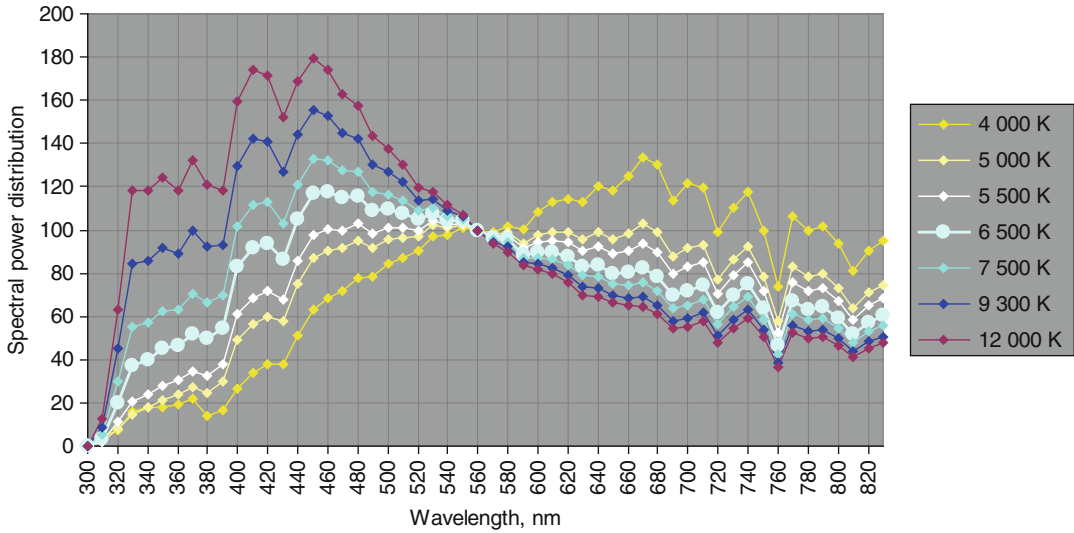
A temperature value in the old scale is to be called *nominal temperature* (T_{nom}). Due to the change of constant c_2 , a temperature value belonging to some nominal temperature is called *actual temperature* (T_{act}) and can be calculated as

$$T_{act} = \frac{1,4388}{1,4380} \cdot T_{nom}. \quad (7)$$

Note that for the calculation of daylight illuminants or reproduction of definition tables of standards or CIE technical reports, the value of the actual correlated color temperature should generally be rounded to 3 decimal places after the decimal sign; see Table 1. If one wants to reproduce the spectra of daylight illuminants, the following hints should be applied besides that given for the rounding of the actual temperature:

Daylight Illuminants, Table 1 Nominal and actual correlated color temperature values of daylight illuminants

Daylight illuminant	Nominal correlated color temperature, K	Actual correlated color temperature, K
D50	5,000	5,002,782
D55	5,500	5,503,060
D65	6,500	6,503,616
D75	7,500	7,504,172



Daylight Illuminants, Fig. 2 Relative spectral power distributions of daylight illuminants belonging to different nominal correlated color temperature values

chromaticity coordinates at Eqs. 4, 5, and 6 have to be rounded to at least 6 decimal places, while factors M_1 and M_2 have to be rounded to exactly 3 decimal places – with the usual way of rounding.

Illustration of daylight illuminants belonging to different nominal correlated color temperature values is presented in Fig. 2.

Interpolation of Daylight Illuminants and Its Consequences

In the 1960s, spectroradiometric methods applied in colorimetry often achieved spectral data with a step size of 10 nm. Daylight illuminants had been defined with the same wavelength step. Later, as spectroradiometric methods improved, more and more demands arose to have daylight illuminants with smaller wavelength steps. As the original measurements could not have been repeated, the only way was to apply interpolation. First, distributions $S_0(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$ were *linearly* interpolated for a wavelength step of 5 nm [4]. CIE recently published a technical report that recommends nonlinear interpolation and gets to smooth spectral distribution curves [5].

In order to get the appropriate spectral power distributions and chromaticity coordinates, some recommendations contain certain rounding rules for computing M_1 and M_2 ; hence, using *other wavelength steps* than the original one of 10 nm, minor differences arise between chromaticity coordinates belonging to correlated color temperature values (x_D , y_D); see Eqs. 4, 5, and 6) and chromaticity coordinates (x_S , y_S) computed from daylight spectral power distributions themselves (from Eq. 1). These differences do not influence practical applications, but they are very embarrassing from a theoretical point of view. This anomaly had not been explained till 2000 [6].

CIE Standard Illuminant D65 (Shortly CIE D65)

As discussed in the previous section, the detailed spectrum of CIE standard illuminant D65 was calculated from the fixed 10 nm values by linear interpolation. Data can be found, e.g., in the book by Wyszecki and Stiles [7] and in the standard for CIE standard illuminant A and D65 [8]. The tabular data were rounded in such a way that 6 significant decimal places were kept, together with the

integer part and the digits of the decimal fraction. It should be noted that an unusual rounding rule had been applied: if a number ended with the digit 5, it was rounded downwards, e.g., at 405 nm the piece of data 87,12045 was rounded to 87,1204.

CIE D65 is one of the most important and most frequently used standard illuminants of the CIE. Many spectrophotometers indicate color values of surfaces in such a way as if they had been illuminated by CIE D65; for further details, see Chapter CIE Standard Illuminants and Sources.

CIE D50, D55, and D75

Although CIE suggests using D65 whenever it is possible, three further illuminants are often used. These illuminants are D50, D55, and D75. D50 is often used in graphic arts; D55 and D75 correspond to phases of daylight when the Sun starts setting and when the blue sky dominates, respectively.

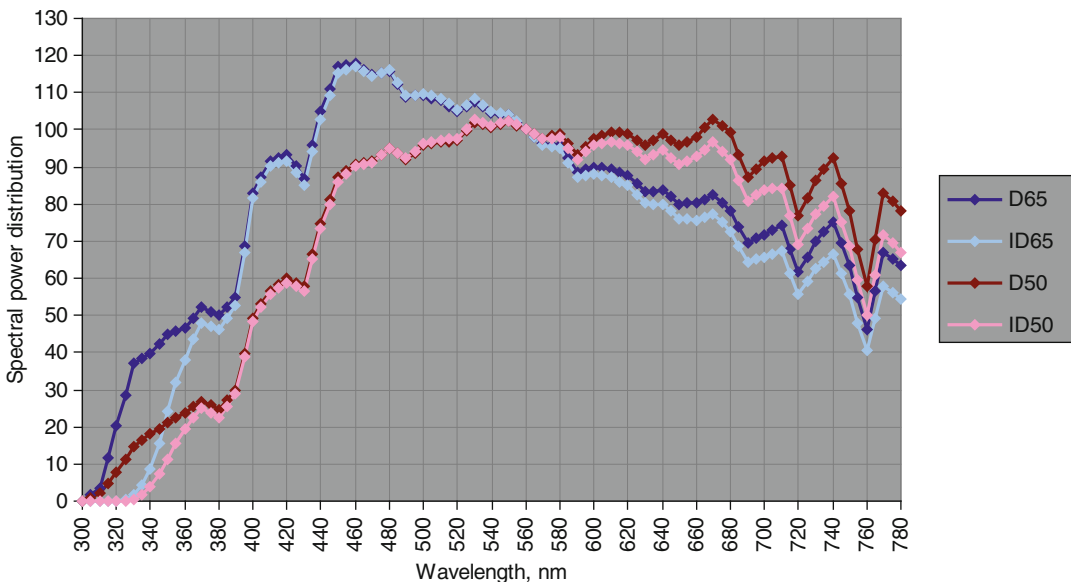
The calculation of D50, D55, and D75 can be performed similarly to that of D65. The definition table in publication CIE 15:2004 [5] contains the wavelength range 300–780 nm with a step size of 5 nm. It should be noted that in this table all

numerical data were rounded to 3 decimal places with the usual rounding rule, independently from the number of digits of the integer part of the values.

Recalculation of Daylight Illuminants

The original CIE rule to calculate the spectra of daylight illuminants is that first the spectrum for the 10 nm steps has to be determined and this spectrum has to be linearly interpolated. If one tries to interpolate first the S_0 , S_1 , and S_2 functions, one will get slightly different values.

It turned out that no rounding rules would give a perfect solution for eliminating the problem of differences between chromaticity coordinates calculated by Eqs. 4, 5, and 6 and from the illuminant itself (Eq. 1). It was found that factors M_1 and M_2 in Eq. 1 depend on the wavelength step of distributions or characteristic vectors S_0 , S_1 , and S_2 . Factors M_1 and M_2 were originally determined for the case of a wavelength step of 10 nm. If any type of interpolation of daylight illuminants is applied, this fact leads to computational inaccuracies and theoretical contradictions.



Daylight Illuminants, Fig. 3 Spectral power distributions of D50, D65, ID50, and ID65

Consequently, factors M_1 and M_2 in Eq. 1 have to be *recalculated* in every case when the wavelength step of distributions S_0 , S_1 , and S_2 differs from 10 nm.

A mathematical procedure was worked out by which factors M_1 and M_2 can precisely be recalculated for any case of interpolation of S_0 , S_1 , and S_2 [6]. Based on the recalculation chromaticity coordinates (x_D , y_D) depending on the correlated color temperature of daylight, distributions always match those, (x_S , y_S), calculated from distribution S . CIE publication 204 contains Excel sheets to perform calculations for any correlated color temperature [5].

Indoor Daylight Illuminants

All daylight illuminants dealt with above have power also in the UV range that cannot be neglected in the case of fluorescent samples. However, natural situations like applying daylight filtered by a window sheet raised a demand for having some indoor versions of daylight illuminants. CIE's technical committee TC 1-66 investigated several plate glass types. Based on their spectral transmittance properties and average thickness, two indoor daylight illuminants were defined: ID50 and ID65 [10]. Figure 3 illustrates D50, D65, ID50, and ID65.

Cross-References

► [CIE Standard Illuminants and Sources](#)

References

1. Judd, D.B., Wyszecki, G., MacAdam, D.L.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
2. Wyszecki, G., Stiles, W. S.: CIE standard illuminants. In: Wyszecki, G., Stiles, W. S. (eds.) *Color Science*, pp. 143–146. Wiley, New York (1982)
3. Schanda, J.: CIE colorimetry. In: Schanda, J. (ed.) *Colorimetry – Understanding the CIE System*, pp. 42–43. Wiley Interscience, Hoboken (2007)

4. Commission Internationale de l'Éclairage, CIE: Colorimetry, 3rd edn, pp. 33–34 CIE **15** CIE Central Bureau, Vienna (2004), Chapter 3.1
5. Commission Internationale de l'Éclairage, CIE: Methods for re-defining CIE D illuminants. CIE **204** CIE Central Bureau, Vienna (2013)
6. Kránicz, B., Schanda, J.: Re-evaluation of daylight spectral distributions. *Color. Res. Appl.* **25**, 250–259 (2000)
7. Wyszecki, G., Stiles, W.S.: Table I(3.3.4). In: Wyszecki, G., Stiles, W.S. (eds.) *Color Science*, pp. 754–758. Wiley, New York (1982)
8. Commission Internationale de l'Éclairage, CIE, Joint ISO/CIE Standard: Colorimetry – Part 2: CIE Standard Illuminants for Colorimetry. CIE S 014-2/E:2006/ISO 11664-2:2007(E) CIE Central Bureau, Vienna
9. Commission Internationale de l'Éclairage, CIE: Colorimetry, 3rd edn, CIE **15** CIE Central Bureau, Vienna (2004), Table T.1, pp. 30–32
10. Commission Internationale de l'Éclairage, CIE: Indoor Daylight Illuminants, 1st edn. CIE **184** CIE Central Bureau, Vienna (2009)

Daylight Spectra

► [Daylighting](#)

Daylighting

John Mardaljevic
School of Architecture, Building and Civil
Engineering, Loughborough University,
Loughborough, Leicestershire, UK

Synonyms

[Correlated color temperatures](#); [Daylight](#); [Daylight spectra](#); [Standard daylight illuminants](#)

Definition

Daylight is the totality of visible radiation originating from the sky and, when visible, the sun during the hours of daytime.

Introduction

The colors or, more specifically, the spectra of the various phases of daylight are closely related to the commonplace experience of naturally occurring sun and sky conditions, e.g., clear blue skies, white clouds, and red sunsets. Prevailing conditions can be characterized in terms of a correlated color temperature. For certain typical sun and sky conditions, there are standardized spectral power distributions. From these it is possible to estimate the time-varying spectra of daylight from illuminance values in standardized climate data files. Thus, it becomes possible to predict the daylight spectra experienced by building occupants. This has found application in studies on the nonvisual effects of daylight and the performance of dynamic glazing systems.

The qualitative aspects of the colors, or hues, of daylight are difficult to describe, harder still to codify in any but a superficial manner. Intricately bound with these perceptions are the conditions under which they occur:

- Prevailing degrees of sunniness and cloudiness
- The position of the sun in the sky (when visible)
- Clearness or haziness of the skies
- Occurrence of various cloud types
- Properties of the natural and man-made reflecting environment

Each of these can have some notion of colors associated with it. How daylight can appear to humans found expression in the world of art where certain qualities of illumination became associated with specific locales, for example, the subtle hues in the gently luminescent skies (often reflected over water) painted by the Dutch masters of the seventeenth century or the bright, vivid shades used by Cezanne and fellow painters who sought inspiration in the sun-drenched south of France. Commonplace holiday photographs can also reveal much of the variety in experienced daylight conditions, both in natural and man-made settings, Fig. 1.

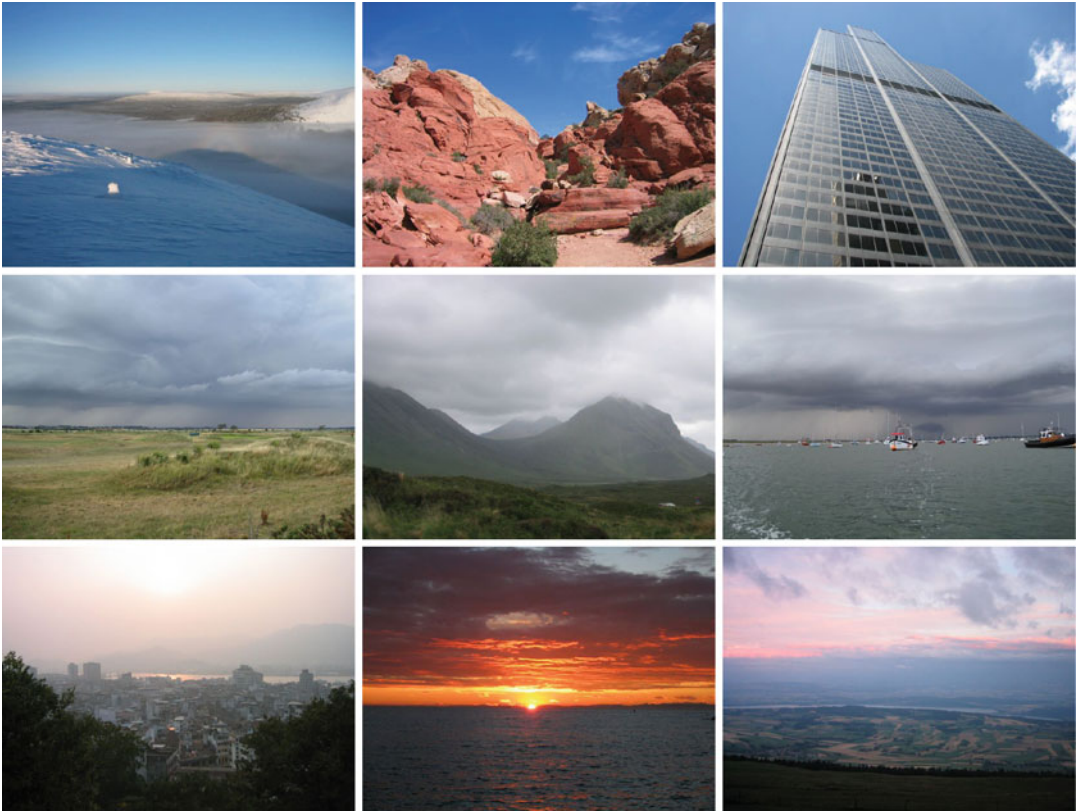
Notwithstanding the evident variety in the hues of the experienced luminous environment,

consideration of the color of daylight has not generally played a role in the daylighting design of buildings – at least not until very recently. In purely aesthetic terms, the color of the daylight is easily noticed, e.g., the warm light from a low sun, the cool light in rooms that are sunless, etc. As a consequence, rooms with a southerly aspect may, in a general rather than just a lighting sense, be perceived as “warm” and those with a northerly aspect as “cool.” This is particularly so in higher-latitude locations, e.g., the Scandinavian countries where the sun is often low in the sky [1]. Designers may even try to compensate somewhat for this, at least for the rooms with a northerly aspect, by selecting paint finishes/decor with warm hues to neutralize somewhat the perception of coolness in the space. There is however little in the way of an empirically based approach to systematically apply such considerations to the color design of spaces.

While it is possible to describe the various types of daylight that occur using chromaticity coordinates [2], it is often simpler and indeed more revealing to consider the spectral power distribution (SPD) of the light rather than reducing it to derived quantities which could hide important details. Recent advances in variable transmission glass technology [3] and in understanding the basis of the nonvisual effects of daylight [4] further suggest that daylight SPDs rather than chromaticity coordinates are needed to progress in these areas. Thus, this chapter will describe daylight and its interactions – reflection and transmission – using spectra rather than any of the color systems. It begins with a consideration of the spectrum of solar radiation arriving at the top of the Earth’s atmosphere.

The Standard Solar Spectrum

The visible radiation is only a part of the total energy associated with daylight radiation. The full spectrum of radiation ranges from about 300 to 4,000 nm, i.e., from the near ultraviolet to the far infrared (both invisible to the eye). The radiation from the sun arriving at the Earth, but before entering the atmosphere, is called extraterrestrial radiation. It approximates a black body radiator with a temperature of 5,800 K, i.e., a



Daylighting, Fig. 1 The “colors” of daylight – light from the sun and/or sky together with that reflected by the environment

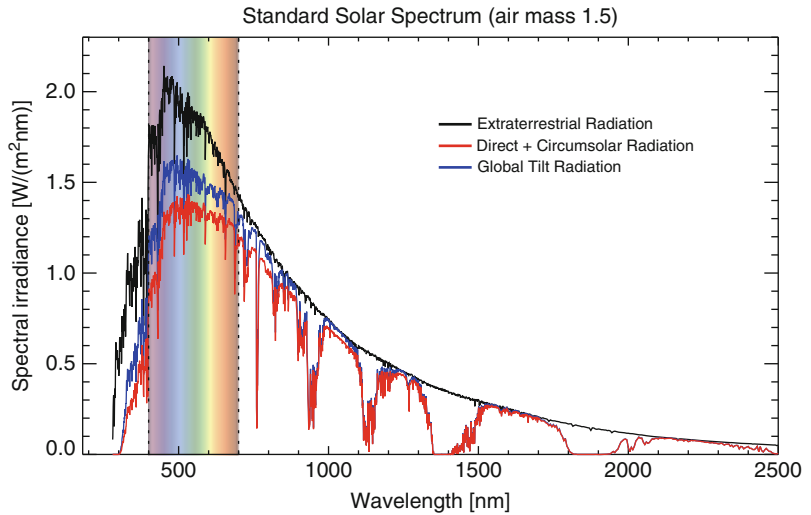
large “tail” at long infrared wavelengths gradually diminishing to zero, the peak of output in the visible spectrum, and a steep cutoff at ultraviolet wavelengths, Fig. 2. How that energy arrives at the surface of the Earth depends on various factors including atmospheric.

Scattering of sunlight in the atmosphere by air, water vapor, dust, etc. gives the sky the appearance of a self-luminous hemispherical source of light. Sunlight is commonly referred to as direct light since it appears to originate from a small source and can cast sharp shadows. The sky however is an extended source of illumination that casts only soft shadows, and so skylight is commonly referred to as diffuse light. Scattering of solar radiation in the atmosphere does not occur equally across the visible spectrum – shorter, i.e., blue wavelength light is scattered more than the

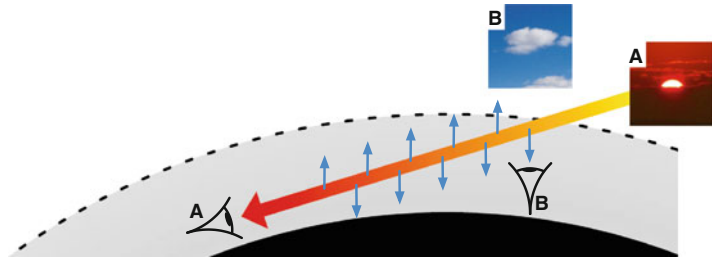
other colors. Thus, on clear days, the sky appears blue – the observer is seeing light (originating from the sun) that arrives via indirect scattered paths. For the same reason, the sun appears yellow because blue light has been scattered out of the beam, i.e., the direct path from the sun to the observer. The lower the angle that the sun appears in the sky, the greater the distance in the atmosphere through which the sunlight travels. Thus, more of the blue (and green) wavelengths are scattered out of the beam, and the sun appears progressively yellow, orange, and then red as the sun sets. A blue sky and a red sunset are in fact the same optical phenomenon – preferential scattering of short wavelength light in the atmosphere – seen from two different viewpoints, Fig. 3. The air mass for a horizontal view (at sea level) is taken to be 38. Thus, the beam of light from the sun at the

Daylighting,

Fig. 2 Standardized extraterrestrial radiation with direct and global radiation received at the surface of the Earth (air mass = 1.5)



Daylighting, Fig. 3 A blue sky and a red sunset – two different views of the same phenomenon



horizon has traveled through approximately 25 times more atmosphere than when seen at an altitude of 41.8° (i.e., the condition for AM1.5).

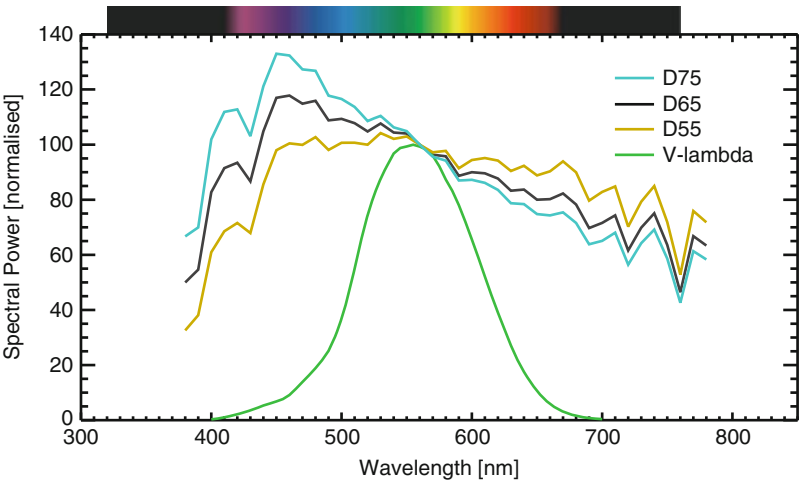
The Standard Daylight Illuminants

The standard solar spectra were largely created for the simulation and testing of energy systems, e.g., photovoltaic panels. Similarly, the standard daylight (or D) illuminants were formulated by the International Commission for Illumination (CIE) to represent various “phases of daylight” [5]. Three of the most commonly used standard D illuminants are D55, D65, and D75. The normalized spectral power distribution (SPD) for each of this is shown in Fig. 4. Also plotted is the visual sensitivity curve $V(\lambda)$ – each curve has been normalized to have the same value at 555 nm. The illuminants are characterized by their nominal correlated color temperature (CCT), i.e., the temperature of the Planckian or blackbody radiator whose perceived color most

closely resembles that of a given stimulus at the same brightness and under specified viewing conditions. The nominal correlated color temperatures for D55, D65, and D75 are, respectively, 5,500 K, 6,500 K, and 7,500 K. The illuminant SPD curves also exhibit the absorption features present in the “direct” and “global” spectral irradiance curves (Fig. 2), though they are now easier to see with the smaller range in plotted wavelength.

In relative terms compared to D65, illuminant D55 will appear “warmer” and D75 “cooler.” Note that higher correlated color temperatures (CCTs) result in SPDs that are perceived as progressively “cooler,” that is, “bluer” even though the (equivalent) blackbody radiator is of course hotter. Of these three, illuminant D65 is taken as the reference source for daylight, corresponding roughly to the illumination from midday sun and (blue) skylight in temperate European locations:

Daylighting,
Fig. 4 Standardized
daylight illuminants and the
visual sensitivity curve



Daylighting, Table 1 Correlated color temperatures for various daylight sources. (Source: Kodak)

Daylight source	CCT
Sunlight – sunrise or sunset	2,000 K
Sunlight – 1 h after sunrise	3,500 K
Sunlight – early morning	4,300 K
Sunlight – late afternoon	4,300 K
Average summer sunlight at noon (Washington, DC)	5,400 K
Direct midsummer sunlight	5,800 K
Overcast sky	6,000 K
Average summer sunlight (plus blue skylight)	6,500 K
Light summer shade	7,100 K
Average summer shade	8,000 K
Summer skylight	9,500–30,000 K

CIE standard illuminant D65 is intended to represent average daylight and has a correlated colour temperature of approximately 6,500 K. [6]

Daylight and Correlated Color Temperature
Data relating natural illumination conditions to correlated color temperatures is sparse. One of the most comprehensive sources of data is the Kodak Cinematographer’s Field Guide [7]. It gives equivalent correlated color temperatures for 11 daylight conditions, Table 1. Here the daylight condition for the nominal CCT of 6,500 K (i.e., D65) is described as “average summer

sunlight (plus blue skylight),” essentially equivalent to the arguably less precise CIE description of the daylight conditions.

Note that the CCT for sunlight varies from 2,000 (sunrise/sunset) to 5,800 K, with “average sunlight at noon” having a CCT of 5,400 K. Thus, illuminant D55 can be assumed to represent the SPD for (direct) sunlight provided that the sun is not too low in the sky. In reality, average summer sunlight plus blue skylight will in fact be a combination of SPDs – D55 for sunlight and some (currently unspecified) SPD for blue skylight with a CCT of up to 30,000 K. The reference illuminant D65 is therefore a compromise offering an SPD that, for most practical purposes, is equivalent to some combination of D55 (sunlight) and some unspecified high CCT for blue skylight. The reference daylight illuminant CCT of 6,500 K is closer to the CCT for average sunlight (5,500 K) than the $\geq 9,500$ K for blue skylight because, in most situations, illumination from the sun (direct and reflected) will dominate that from the sky.

Kodak recommends that an overcast sky has an SPD equivalent to a CCT of 6,000 K. In which case, it would appear that the closest match of the three D illuminants is either D55 or D65. Given that D55 is believed to be a good match for sunlight, the SPD for an overcast sky is better represented by D65. That has been confirmed by measurements taken on a wide range of commonly occurring overcast sky conditions [8].

Daylight Quantities in the Standard Climate Datasets

The quantitative nature of illumination from the sun and sky for any particular locale can be appreciated by examining the luminous output from both the sun and the sky over a period of a full year. A full year is needed to capture seasonal variation in daylight. The principal sources of annual data for daylight are the standard weather files which were originally created for use by dynamic thermal modeling programs [9]. These datasets contain averaged hourly values for a full year, that is, 8,760 values for each parameter. The key daylight parameters stored in the weather files are the diffuse horizontal illuminance and the direct normal illuminance. The diffuse horizontal illuminance is the visible light energy from the unobstructed sky that is incident on a horizontal surface. The direct normal illuminance is the visible light energy from the sun that is incident on a surface which is kept normal to the beam of radiation, i.e., the photocell always “points” directly to the sun. The illuminating power of the sun is measured in this way because when the sun is low its contribution to horizontal illuminance is small even if the absolute intensity is high.

A visualization of the illuminance data from the standard weather files for Madrid (Spain) and Ostersund (Sweden) is given in Fig. 5. The time-series data of 8,760 values has been rearranged into an array of 365 days (x-axis) by 24 h (y-axis). The shading at each hour indicates the magnitude of the illuminance – see legend – with zero values shaded gray. Presented in this way, it is easy to appreciate both prevailing patterns in either quantity or their short-term variability. Most obvious is the daily/seasonal pattern for both illuminances: short periods of daylight in the winter months and longer in summer. The hour-by-hour variation in the direct normal illuminance is clearly visible, though it is also present to a lesser degree in the diffuse horizontal illuminance (i.e., light from the sky).

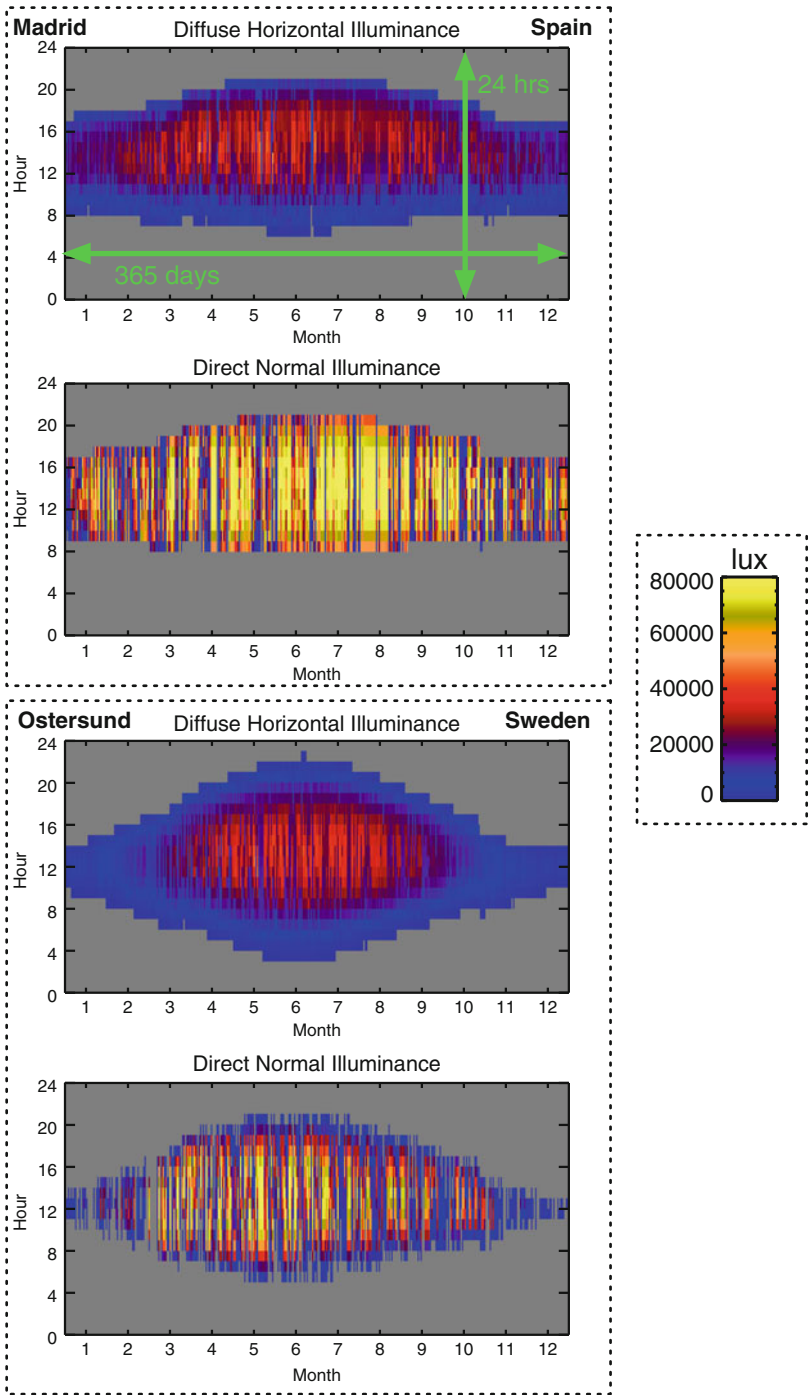
Comparing the data for the two locales, one immediately notices the pronounced seasonal variation in day length for Ostersund, which is in the north of Sweden. As might be expected, the Madrid data show many sunny days, especially

in summer. Note that Spain uses Central European Time despite having a longitude similar to the UK. Thus, there is a marked divergence between clock and solar time in the Madrid data. The diffuse horizontal illuminance reaches a maximum of about 40,000 lx (red shading), and direct normal illuminance under clear sky conditions often exceeds 70,000 lx (yellow shading). This is so for both locales. During winter in northern Sweden, the diffuse horizontal illuminance is typically 5,000 lx or even less, with few hours of sun.

Both diffuse and direct illuminances will, in reality, vary over periods much shorter than an hour. Interpolation of the dataset to a time-step shorter than 1 h will provide a smoother traversal of the sun, which may be necessary when using the data for simulation of daylight. Interpolation alone however will not introduce short-term variability into the values for diffuse horizontal and direct normal illuminance. The patterns of hourly values in the illuminance datasets are unique, and because of the random nature of weather, they will never be repeated in precisely the way shown in Fig. 5. Climate datasets are however representative of the prevailing conditions measured at the site, and they do exhibit much of the full range in variation that typically occurs, i.e., they provide definitive yardstick quantities for modeling purposes. These datasets are used to define the external conditions for climate-based daylight modeling (CBDM), i.e., the prediction of annual profiles of internal (daylight) illuminance due to all the unique sun and sky conditions derived from a standard climate file [10].

For every value in the datasets for diffuse horizontal and direct normal illuminance, it is possible to estimate, with varying degrees of reliability, a corresponding CCT using the entries in Table 1. For example, much of the intense sunlight (i.e., direct normal illuminance) could be taken to have a CCT of around 5,400–5,800 K. Thus, the SPD for the sunlight could be approximated to standard illuminant D55. Sunlight around the hours of dawn and dusk could be assigned lower CCTs, e.g., 2,000–3,500 K, while early-morning and late-afternoon sun could be approximated to a source with a CCT of 4,300 K. Use of climate files to estimate the SPD of the sun and skylight

Daylighting,
Fig. 5 Illuminance data
from standard climate files
for Madrid (Spain) and
Ostersund (Sweden)



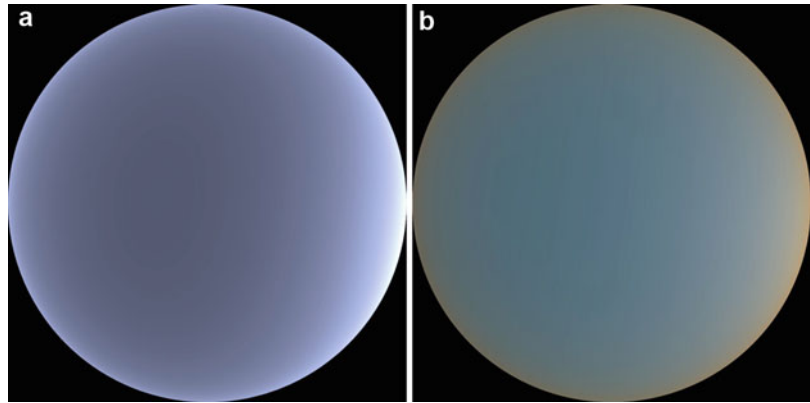
sources of illumination has recently been demon-
strated for the purpose of modeling the nonvisual
effects of daylight (see section “[The Nonvisual
Effects of Daylight](#)”).

Sky Models

A sky model is a mathematical equation that
describes the variation in luminance across the
hemispherical sky vault. The simplest possible

Daylighting,

Fig. 6 Simulated 180° hemispherical “fish-eye” images of a constant chromaticity clear sky distribution (a) and a variable chromaticity clear sky (b)



model describes a sky of uniform luminance, i.e., the same brightness is seen wherever one looks above the horizon (in an unobstructed setting). At the start of the twentieth century, it was believed that overcast skies exhibited only moderate variation in brightness across the sky dome, and so they could be considered to be of constant (i.e., uniform) luminance. Measurements revealed however that a densely overcast sky exhibits a relative gradation from darker horizon to brighter zenith; this was recorded in 1901. With improved, more sensitive measuring apparatus, it was shown that the zenith luminance is often three times greater than the horizon luminance for some of the most heavily overcast skies [11]. The formulation for the luminance pattern of overcast skies presented by Moon and Spencer in 1942 was adopted as a standard by the International Commission for Illumination (CIE) in 1955.

The equations describing the luminance pattern for clear or intermediate skies are rather more complex than that for the CIE standard overcast sky. The complexity arises from a number of observed effects that are accounted for in the model. Among these are a bright circumsolar region, a sky luminance minimum that is at some point above the horizon, and a brightening of the sky near the horizon. The scales of these effects are related to the solar position and the relative magnitudes of the illumination produced by the sun and sky [12]. These and similar sky models, e.g., the Perez all-weather model [13], do not include provision in the formulation for any variation in chromaticity of the luminance across

the sky vault. If the sky color is required in a daylight simulation, the RGB radiance values applied to the self-luminous “material” used to model the sky can be set to achieve the desired effect, e.g., a small blue “excess” to create a blue hue for the clear sky distribution, Fig. 6a. Notwithstanding the variation in luminance across the sky vault, the chromaticity will remain constant.

More recently, sky models with varying chromaticity have been formulated to depict the subtle color gradations seen in nature [14]. Based on analytical models for the scattering of light in the atmosphere, these formulations can account for effects such as blue sky at high altitude together with warm (i.e., blue deficient) sky just above the horizon, Fig. 6b. It should be noted that data to validate these models is scarce, and, for daylight simulation, their use is primarily to lend realism to renderings rather than for any quantitative modeling, e.g., illuminance prediction.

Daylight in Buildings

Daylight illumination of a space can be considered as a series of transforms – from a light emitted by the luminous source(s) to that received at a point in a space. For daylight performance evaluation, practitioners have largely been concerned with the amount of light received at a point in a space (i.e., lux levels) and have not been concerned with the spectral nature of any of the transforms. For most quantitative modeling purposes – using either the long-standing daylight

factor method or more recent climate-based daylight modeling – it is sufficient to describe all opaque surfaces as gray reflectors. And, although tinted glazing materials may be represented in the simulation by their RGB spectral transmittance values, an equivalent gray transmittance would work just as well.

Color can, of course, be modeled in a lighting simulation. Though, depending on the tool used, there may be limitations or restrictions to the degree of precision that can be achieved when modeling the spectral properties of reflection and transmission. Key among these are the limitations of having three channels, as is the case with *Radiance* – generally considered the most widely used and rigorously validated lighting simulation package [15]. The “color” assigned to each of the three channels is arbitrary, though in most instances they are used to model the standard RGB trichromatic system common in computer graphics and digital image display technology. While adequate for many if not most image display purposes, the three channel RGB approaches are in fact a fairly coarse representation of the reflection and transmission properties of materials. For example, it is possible to have materials with (noticeably) different spectral reflectance curves giving the same RGB values [16]. Thus, although noticeably different to an observer, they would all have the same representation in the simulation, i.e., the same RGB spectral reflectance values.

Transformation of Daylight by Reflection and Transmission

Considering only the normalized spectra (i.e., ignoring absolute values), one can represent the transformations fairly simply as follows. Consider a space with a window receiving direct sun on sunny day with a clear blue sky. The SPDs for the sunlight and the skylight incident on the window are denoted, respectively, by the vectors $\mathbf{D}^{\text{sun}}$ and $\mathbf{D}^{\text{sky}}$. The SPD for the light received at the window will be comprised of these two illuminants plus sunlight and skylight reflected off the external surfaces, e.g., the ground and nearby buildings. The vector for the reflected SPD incident on the window is denoted by $\mathbf{D}^{\text{ref}}$. The reflected SPD will be some combination of the SPDs from the

sky and (if present) the sun, modified by the spectral reflectance properties of the ground, external buildings, etc. The SPD for the light incident at the window $\mathbf{D}^{\text{icd}}$ is given by:

$$\mathbf{D}^{\text{icd}} = \mathbf{D}^{\text{sun}} + \mathbf{D}^{\text{sky}} + \mathbf{D}^{\text{ref}} \quad (1)$$

If, say, the window is facing the sun direction, then of the total light incident on the glass that arriving directly from the sun will probably form the greatest part. And so the incident SPD could be approximated to the D55 illuminant. As noted previously, only the shape of the spectra is important for this illustration.

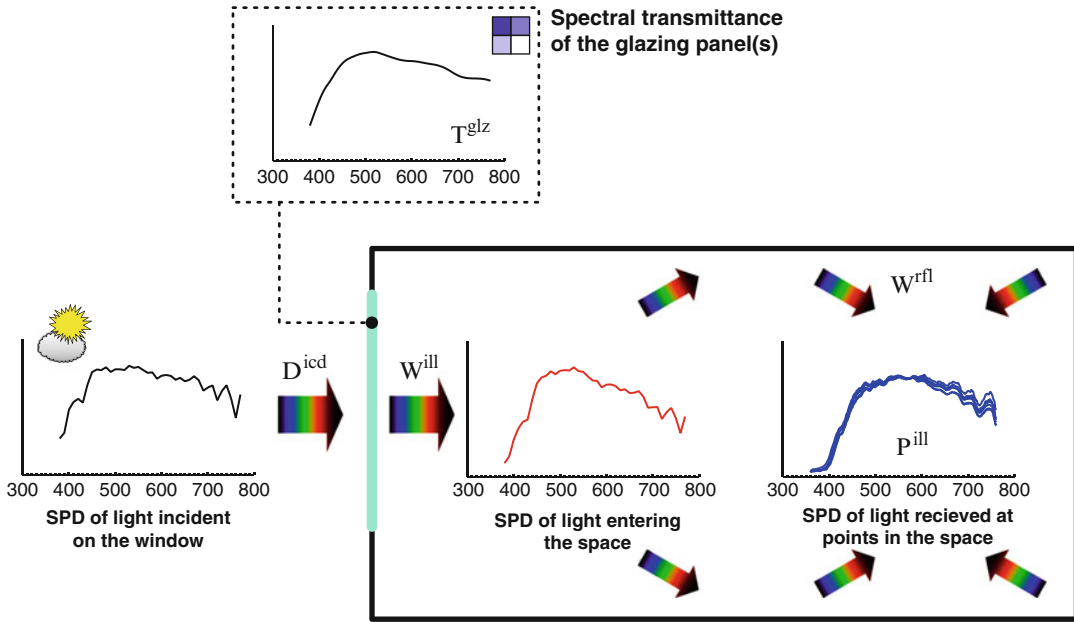
The spectrum of light incident on the window will, on passing through it, be modified by the spectral transmittance $\mathbf{T}^{\text{glz}}$ of the glazing material. The element-by-element (or Hadamard) product of the incident SPD and spectral transmittance vectors give the SPD for the light entering the space:

$$\mathbf{W}^{\text{ill}} = \mathbf{D}^{\text{icd}} \circ \mathbf{T}^{\text{glz}} \quad (2)$$

The vector $\mathbf{W}^{\text{ill}}$ may be called the window illuminant SPD since this is the daylight from the window illuminating the space. This light can arrive at a point in the space directly (if the point can “see” the window) and indirectly following one or more reflections. The SPD of the window illuminant will in turn be transformed following every possible reflection in the space depending on the spectral reflectance properties of the various surfaces. Thus, for every point in the space (x, y, z, θ, ϕ) , the light falling on it $\mathbf{P}^{\text{ill}}$ will have a unique SPD comprised of direct and reflected SPDs:

$$\begin{aligned} \mathbf{P}^{\text{ill}}(x, y, z, \theta, \phi) &= \mathbf{W}^{\text{ill}}(x, y, z, \theta, \phi) \\ &+ \mathbf{W}^{\text{ref}}(x, y, z, \theta, \phi) \end{aligned} \quad (3)$$

In addition to the (x, y, z) position in the space, the orientation of the receiving surface (θ, ϕ) needs to be defined also – the “surface” could be the eye of an observer. If the surface cannot “see” the window directly, then the $\mathbf{W}^{\text{ill}}(x, y, z, \theta, \phi)$ term will be zero. This process of spectral transformations is illustrated in Fig. 7.



Daylighting, Fig. 7 Transformations in the SPD from source(s) to a point inside a space

Practical Simplifications

Under many actually occurring conditions, the SPD of (visible) radiation incident on the windows of a building in a man-made setting can be assumed to be dominated by light from the sky and sun (if present). Thus, one can neglect the transform in the SPD from the reflected component because (a) the ground reflectivity is invariably low and (b) man-made external surfaces in any case tend to be fairly neutral in color. Under these conditions, it is reasonable to assume that the SPD incident on the windows will be approximately that of the dominant source. For example, if direct sun is incident on the window of a building, we can assume that the incident SPD is that of the standard illuminant D55, i.e., $D^{icd} = D^{55}$.

In an actual building with typical neutral decor, the SPD is likely to vary only marginally with position in the space. In other words, inside the space the window illuminant SPD is not transformed significantly on reflection. One can make the approximation that $W^{rf} \simeq W^{ill}$ since the variation with position (x, y, z, θ, ϕ) in the space is not a significant factor. Thus, one can assume that the (normalized) SPD at any point in the space is

approximately equal to the SPD of the light entering the space through the window, i.e., $P^{ill} \simeq W^{ill}$.

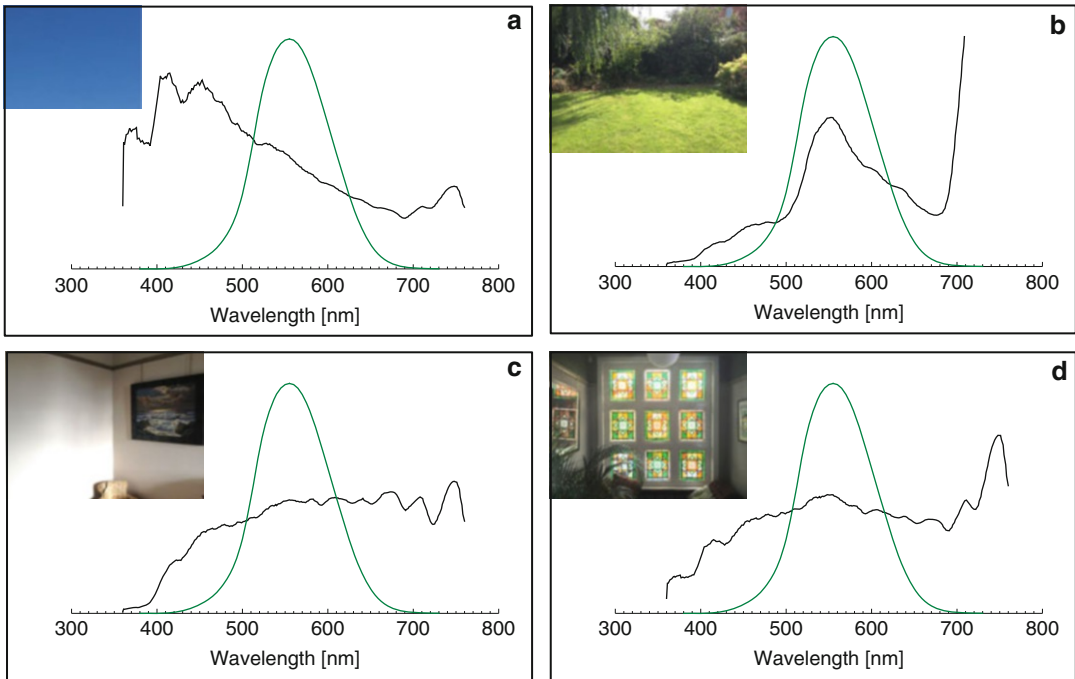
Combining these two simplifications, the SPD of light in the space (P^{ill}) is now governed by just two factors: the SPD of the incident light and the spectral transmittance of the glazing. Using the above example with direct sun incident on the window, the equation for the SPD in the space simplifies to:

$$P^{ill} \simeq D^{55} T^{glz} \quad (4)$$

This approximation has been tested and shown to be perfectly adequate for the purpose of predicting the SPD in a space (see section “[Variable Transmission Glazing](#)”).

Measured Daylight Spectra

It is now possible to take reliable measurements of SPDs using relatively inexpensive, handheld spectrometers. These devices measure spectral irradiance, which is approximately that seen by the eye when “pointing” in the same direction. Four measured daylight SPDs are shown in Fig. 8. The inset



Daylighting, Fig. 8 Measured daylight spectra: (a) blue sky, (b) sunlit garden, (c) back of a room with warm/neutral decor and clear glazing on a bright day, and, (d) daylight from a stained glass window

photograph shows approximately what was “seen” by the sensor, and the visual sensitivity curve $V(\lambda)$ is also shown. The SPDs were normalized against $V(\lambda)$ to have the same area in each case, i.e., to have the same lux value irrespective of the shape of the spectra. This allows a like-for-like comparison of the SPDs whatever their absolute values are.

Two Outdoor Daylight Spectra

The SPD for light from the bluest part of a clear sky shows pronounced peaks around 400 nm, Fig. 8a. Comparison with the standard illuminants (Fig. 4) indicates some similarity on the basic form; however, this light would have a CCT significantly higher than the 7,500 K for $D75$ (the spectrometer reported a CCT of over 12,000 K for this SPD). The SPD for the sunlit garden shows a strong peak in the green part of the spectrum due to reflection from the foliage and a steep rise at around 700 nm at the start of the infrared region, Fig. 8b.

Two Indoor Daylight Spectra

The SPD taken “looking” at the back of a daylighted room with warm/neutral decor and clear glass windows is shown in Fig. 8c. The measurement was taken on a sunny day, and, although there was no direct sun in the space at that time, the SPD incident on the window would probably have been dominated by reflected sunlight. The decor in the space would be described as warm/neutral since there are no highly saturated colors present. The SPD is fairly flat over most of the visible range, and the illumination in the space would indeed be described as essentially neutral. In the range of 600–700 nm, it is possible to discern some of the features present in the standard illuminants (Fig. 4). Note however that the SPD for daylight indoors cannot in the main correspond to $D65$ because all standard glass materials absorb light at short wavelengths.

The SPD for the light from the stained glass window is perhaps rather more neutral than one might expect from the photograph of the window, Fig. 8d. The visual impression of the window

suggests that greens and yellows will be dominant in the SPD. However, for the window as a whole, the bulk of the light transmission occurs through the clear sections of the glass, resulting in a fairly neutral spectrum overall. This is because the tinted glass has a lower transmission than clear glass. The more saturated the color of the stained glass, the lower its transmission, and so the smaller its contribution to overall illumination. Notwithstanding the striking visual impression made by stained glass windows, the overall SPD of the light through them can be surprisingly neutral provided that a proportion of the glass in the window is clear.

Applications

Perhaps the first practical application of the science of daylight color was in the field of cinematography, where the various phases of daylight were associated with a correlated color temperature (see Table 1). In moviemaking, one scene could take several days to film, and if done so under natural lighting, it is important to maintain color consistency between the various “takes.” Cameramen can use subtle filters, or film stock with different spectral sensitivities, to help maintain a desired CCT between “takes” filmed on days with different daylight conditions [7]. Filming indoors was largely carried out under artificial lighting. Thus, any consideration of the color of daylight indoors was largely an aesthetic concern [1].

However, recent advances in:

- (i) understanding the basis for the non-visual effects of daylight; and,
- (ii) the manufacture of variable transmission glazing,

have caused scientists and practitioners to pay attention to the color, or more specifically, the spectra of daylight experienced indoors by the occupants of buildings. The sections below describe some of the first studies to model (section “[The Nonvisual Effects of Daylight](#)”) and measure (section “[Variable Transmission](#)

[Glazing](#)”) daylight spectra for the two application scenarios noted above.

The Nonvisual Effects of Daylight

It is now well recognized that illumination received at the eye is responsible for a number of effects on the human body that are unrelated in any direct sense to vision. Light has measurable neuroendocrine and neurobehavioral effects on the human body, in particular with respect to maintaining a regular sleep-wake cycle that is entrained to the natural diurnal cycle of night and day [17]. Additionally there is evidence to suggest links between daylight illumination in particular and alertness, productivity, and academic achievement [18]. In the early 1990s, it was discovered that the eye possesses an additional non-rod, non-cone photoreceptor named photosensitive retinal ganglion cells or pRGCs [19]. The pRGCs communicate light-induced signals to the suprachiasmatic nuclei (sometimes referred to as the “circadian pacemaker”) in the anterior hypothalamus of the brain. This and related discoveries have led to the emergence of a new area for photobiology research concerned with daylight/light exposure and its nonvisual effects [20].

There are numerous factors operating in a building that influence the sense of well-being or, otherwise, that an occupant experiences. Daylight is known to be one of them. The emergence of the field of photobiology has led to the speculation that we might one day be able to predict for an interior space a measure of the nonvisual effect of the daylight provided by the design of the building. This speculative measure has been referred to as the “circadian potential” or “circadian efficiency” of the space [21].

The spectral sensitivity of the photosensitive retinal ganglion cells is shifted to the blue compared to the visual sensitivity $V(\lambda)$ curve. The precise shape is not known, but the peak of sensitivity is around 460–480 nm. The timing, intensity, spectrum, duration, pattern of light received at the eye, and prior light history are the principal factors determining entrainment of the circadian cycle. Although the understanding of the various factors, and their interrelations, is incomplete,

there is now sufficient empirical data to begin to experiment with predictive models for the non-visual effects of light – artificial as well as daylight [22]. With regard to the spectrum of light received at the eye, for two sources of illumination delivering the same lux value, one that has a blue-rich SPD could produce a stronger nonvisual effect than one which has less power in the 460–480 nm region [23]. Thus, any materials that significantly transform the spectrum of daylight entering a space should, once the scientific basis is sufficiently refined, be assessed for their potential to maintain desired nonvisual effects in addition to providing illumination for task.

Variable Transmission Glazing

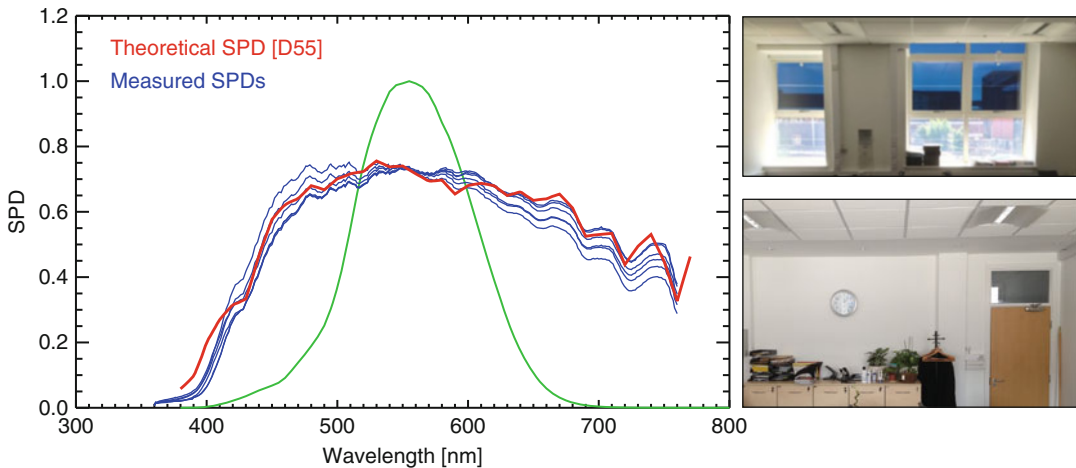
The use of daylight in office buildings is generally considered to be a greatly underexploited resource. In large part this is because of the highly variable nature of daylight illumination. The natural, large variability in daylight means that users will often need to use shades to moderate excessive ingress of daylight. Most shading systems act as a “shutter” that is either open or closed, with users rarely making the effort to optimize the shading for both daylight provision and solar/glare control. And blinds are often left closed long after the external condition has changed. A glazing with a transmissivity that varies continuously between clear and dark extremes could offer a much greater degree of control over the luminous environment.

The principle behind variable transmission glazing (VTG) is straightforward: the transmission properties of the glazing are varied to achieve an “optimum” luminous and/or thermal environment. In practice, developing the technology has proven challenging. Of the various VTG formulations undergoing research and development over the last two decades, the most promising appears to be electrochromic where a small applied voltage varies the transmission properties of the glass [24]. One of the most advanced of these materials is produced by SAGE Electrochromics Inc., and in 2013, it started being manufactured in high volume. This EC glass varies in visible transmission from 60% when in the clear state down to 1% (or even less) when fully tinted. As the glass

darkens (i.e., “tints”), the longer wavelengths are diminished proportionally to a greater degree than the shorter wavelengths, giving the EC glazing good solar control properties to help prevent overheating. Optically, the consequence of this is to shift the peak in visible transmission to the blue end of the spectrum. However, while control of the luminous and thermal environment is highly desirable, occupants are believed to prefer daylight illumination that is perceived as neutral rather than tinted. Thus, the question regarding the neutrality of the illumination spectrum is an important one that needs to be addressed.

Effectively neutral illumination in spaces with EC glass can be achieved provided that a proportion of the glass remains in the clear state [25]. This was demonstrated by taking measurements of the SPD of ambient illumination in a daylighted office with EC glass under sunny conditions. The measurements were compared with a theoretical model for the predicted SPD. The theoretical SPD for the illumination in the space was predicted using the simplified expression in Eq. 4. Of the eight EC panels in the window of the office used for the evaluation, five were set to full tint and three were in the clear state [26]. The spectral transmittance of the eight glass panels in the various states was calculated as a composite spectral transmittance curve based on the individual spectral transmittance curves and their proportions. Thus, in terms of the SPD, the eight panels in the various states were modeled as a single entity.

The measured SPDs were taken at six points in the office: four corresponding to the view positions of the four occupants who normally work in this space, one at the back of the office looking toward the window, and one at the window looking toward the back of the space. The six measured SPDs (normalized as before) are shown in Fig. 9. Notwithstanding the fact that the SPDs were measured at different points across the room – both looking toward and away from the window – the SPDs are in fact very similar. Furthermore, the SPDs are fairly neutral across the visual range, and the $V(\lambda)$ curve is also shown. This is consistent with experience of such spaces where the illumination turns out to be rather more neutral than one might expect from



Daylighting, Fig. 9 Theoretical and measured SPDs for an office with electrochromic glazing – with photographs of the EC window (*upper*) and the rear of the space (*lower*)

the visual impression of the window – provided that a proportion of the glass is clear. Compare this with the illumination SPD for daylight through the stained glass window (Fig. 8d), where a similar effect is observed. Also shown in Fig. 9 is the predicted SPD calculated using the D55 illuminant and a spectral transmittance properties of the composite [27]. The agreement is remarkably good, and, while these are results from a single study, they provide encouragement for wider and deeper investigations of daylight SPDs experienced by the occupants of buildings.

The Future

The daylighted indoor luminous environment can be quite different to that experienced outdoors. Appreciation of the recently discovered nonvisual effects of illumination has led to a reconsideration of the function of daylight in buildings. The quantification and evaluation of the spectral nature of daylight experienced by building occupants are now achievable using fairly modest apparatus. In the immediate future, it is likely that there will be efforts to characterize the daylight experienced indoors in terms of both the spectrum received at the eye and also the absolute (visual) magnitude of

the illumination delivered for task. When the relation between daylight and nonvisual effects becomes better understood, practitioners will then have a basis to refine the appreciation of what constitutes “sufficiency” for daylight provision. That may have far-reaching consequences for the drafting of future standards for daylight in buildings.

Cross-References

- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)

References

1. Hårleman, M., Werner, I.-B., Billger, M.: Significance of colour on room character: study on dominantly reddish and greenish colours in north- and south-facing rooms. *J. Int. Colour. Assoc.* **1**, (2007)
2. Hernández-Andrés, J., Romero, J., Nieves, J.L., Lee Jr., R.L.: Color and spectral analysis of daylight in southern Europe. *J. Opt. Soc. Am. A.* **18**(6), 1325–1335 (2001)
3. Sanders, H., Andreau, A.: The next generation of dynamic glazing product performance. In: *Proceedings of Glass Performance Days*, Tampere, 13–15 June 2013

4. Webb, A.R.: Considerations for lighting in the built environment: non-visual effects of light. *Energy Build.* **38**(7), 721–727 (2006)
5. Commission Internationale de l'Eclairage: Colorimetry. CIE 15-2004 (2004)
6. Commission Internationale de l'Eclairage: Colorimetry – Part 2: CIE Standard Illuminants. 014-2E:2006 (2006)
7. Kodak: Cinematographer's Field Guide, 14th edn. (2010)
8. Lee Jr., R.L., Hernandez-Andrés, J.: Colors of the daytime overcast sky. *Appl. Opt.* **44**(27), 5712–5722 (2005)
9. Clarke, J.A.: *Energy Simulation in Building Design*, 2nd edn. Butterworth-Heinemann, Oxford (2001)
10. Mardaljevic, J., Heschong, L., Lee, E.: Daylight metrics and energy savings. *Light. Res. Technol.* **41**(3), 261–283 (2009)
11. Moon, P., Spencer, D.E.: Illuminations from a non-uniform sky. *Illum. Eng.* **37**, 707–726 (1942)
12. Commission Internationale de l'Eclairage: Standardization of Luminance Distribution on Clear Skies. CIE 22 (TC-4.2) (1973)
13. Perez, R., Seals, R., Michalsky, J.: All-weather model for sky luminance distribution-preliminary configuration and validation. *Sol. Energy.* **50**(3), 235–245 (1993)
14. Preetham, A.J., Shirley, P., Smits, B.: A practical analytic model for daylight. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH' 99*, pp. 91–100. ACM Press/Addison-Wesley, New York (1999)
15. Ward Larson, G., Shakespeare, R., Mardaljevic, J., Ehrlich, C., Phillips, E., Apian-Bennewitz, P.: *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaufmann, San Francisco (1998)
16. Schanda, J. (ed.): *Colorimetry: Understanding the CIE System*. Wiley, Hoboken (2007)
17. Lockley, S.W., Dijk, D.J.: Functional genomics of sleep and circadian rhythm. *J. Appl. Physiol.* **92**, 852–862 (2002)
18. Heschong, L.: Daylighting and human performance. *ASHRAE J.* **44**(6), 65–67 (2002)
19. Foster, R.G., Provencio, I., Hudson, D., Fiske, S., De Grip, W., Menaker, M.: Circadian photoreception in the retinally degenerate mouse (rd/rd). *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* **169**(1), 39 (1991)
20. Brainard, G.C., Hanifin, J.P.: Photons, clocks and consciousness. *Biol. Rhythm.* **4**(20), 314–325 (2005)
21. Pechacek, C.S., Andersen, M., Lockley, S.W.: Preliminary method for prospective analysis of the circadian efficacy of (day) light with applications to healthcare architecture. *Leukos.* **5**(1), 1–26 (2008)
22. Andersen, M., Mardaljevic, J., Lockley, S.W.: A framework for predicting the non-visual effects of daylight – part I: photobiology-based model. *Light. Res. Technol.* **44**(1), 37–53 (2012)
23. Brainard, G.C., Hanifin, J.P., Greeson, J.M., Byrne, B., Glickman, G., Gerner, E., Rollag, M.D.: Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**(16), 6405–6412 (2001)
24. Mortimer, R.J.: Electrochromic materials. *Annu. Rev. Mater. Res.* **41**(1), 241–268 (2011)
25. Mardaljevic, J.: Variable transmission glazing: a viable retrofit technology? In: *5th VELUX Daylight Symposium, Copenhagen*, 4–5 May 2013
26. Kelly, R., Painter, B., Mardaljevic, J., Irvine, K.: Capturing the user experience of electrochromic glazing in an open plan office. In: *CIE Midterm Conference – Towards a New Century of Light, Paris*, 12–19 Apr 2013
27. J. Mardaljevic, R., Kelly Waskett., Painter, B.: Neutral daylight illumination with variable transmission glass: Theory and validation. *Lighting Research and Technology*, **48**(3), 267–285 (2016)

d-Block Elements

- [Transition Metal Ion Colors](#)

DE*ab

- [CIELAB](#)

Deep Compositing

- [Compositing and Chroma Keying](#)

ΔE^*94

- [CIE94, History, Use, and Performance](#)

$\Delta E00$

- [CIEDE2000, History, Use, and Performance](#)

Demichel Equations

Roger D. Hersch

Ecole Polytechnique Fédérale de Lausanne
(EPFL), Lausanne, Switzerland

Synonyms

Halftone layer superposition equations

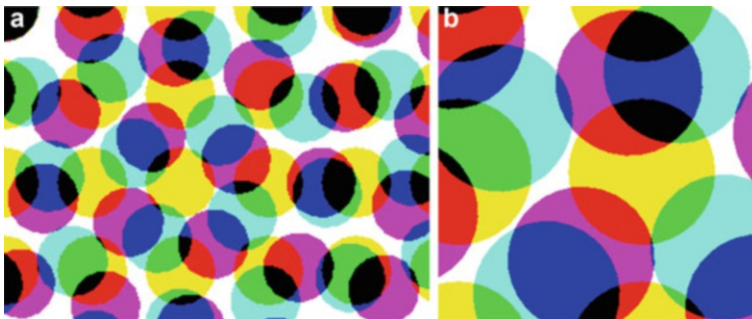
Definition

In color halftoning, the Demichel equations define the relationships between the surface coverages of the printed ink dots and the surface coverages of the Neugebauer primaries, i.e., the surface coverages of the colorants populating the color image. These equations are valid for the independently laid out ink layers that occur in stochastic screening, in error diffusion halftoning, and in mutually rotated clustered dot or line screens. The Demichel equations are used in connection with spectral models predicting the reflectances of color prints in function of the deposited amount of inks [1, 2].

Overview

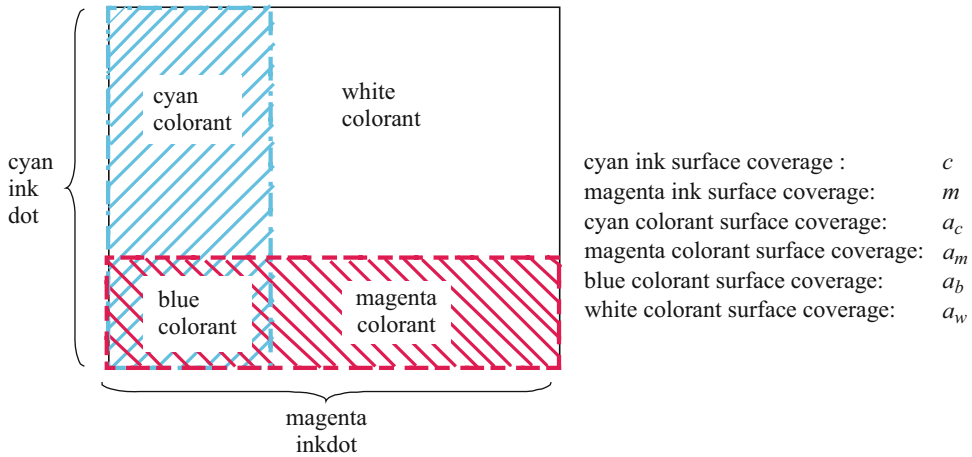
Professional as well as desktop printers are either able to print with a limited number of pixel dot sizes or with a single pixel dot size. For example, in modern ink-jet printers, several droplets may be printed at the same pixel location, leading to a limited number of pixel dot sizes. In classical offset prints, a pixel location is either printed at a fixed size or remains unprinted. Color variations are created by halftoning, i.e., by creating with the different colored inks halftone dots having predetermined surface coverages.

For the sake of clarity, let us consider a cyan, magenta, and yellow ink print halftoned with classical mutually rotated halftone dots. As shown in Fig. 1, the superposition of ink dots of different colors yields new colors. The superposition of a cyan and a yellow ink dot yields a green dot portion; the superposition of a cyan and a magenta ink dot yields a blue dot portion; the superposition of a magenta and a yellow ink dot yields a red dot portion; and the superposition of a cyan, a magenta, and a yellow dot yields a black dot portion. Like in a puzzle, the space of a color halftone comprises juxtaposed dot portions of the colors white (no ink), cyan (cyan only), magenta (magenta only), yellow (yellow only), red (magenta superposed with yellow), green (cyan



Demichel Equations, Fig. 1 (a) Clustered dot color halftone with the *cyan* layer at orientation 15°, the *magenta* layer at orientation 75°, and the *yellow* layer at orientation 0° and (b) enlarged part of the halftone showing the new

colorants *red*, *green*, *blue* in regions where two of the *cyan*, *magenta*, and *yellow* ink dots overlap and *black* where the three ink dots overlap



Demichel Equations, Fig. 2 Example of a halftone with one *cyan* ink halftone dot (*dashed-dotted*) and one *magenta* ink halftone dot (*dashed*), yielding four different colorant surfaces, namely *white*, *cyan* only, *magenta* only, and *blue*

superposed with yellow), blue (cyan superposed with magenta), and black (cyan superposed with magenta and yellow). These colored dot portions are called “Neugebauer primaries” or simply “colorants.”

The *Demichel equations* enable calculating the relative surface covered by each colorant. They assume that the ink layers are printed independently one from another. In the present context, independence means that if a light ray hits a given position on the paper, the probabilities that the different halftone ink layers contain an ink dot at this position are independent one from another. As illustration, consider a first layer (cyan) whose halftone dot grows horizontally and a second ink layer (magenta) whose halftone dot grows vertically.

Let us assume that the total surface of the halftone element is one. The respective surface coverages of the cyan and magenta inks are respectively c and m , having values between 0 and 1. Let us assume that light rays are thrown into the halftone element surface, and that their probability to fall within a specific location of that surface follows a uniform distribution. Then the probability $P(c)$ to fall on a cyan ink dot of surface coverage c is equal to c and the probability $P(m)$ to fall on a magenta ink dot of surface coverage m is equal to m .

The halftone surface is a “puzzle” of juxtaposed colorants, formed by the paper, the inks, and their superpositions. In Fig. 2, the *colorant cyan* is the region covered by ink cyan only, without superposed magenta, and the *colorant magenta* is the region covered by ink magenta only, without superposed cyan. *Colorant white* is the region covered by no ink, and *colorant blue* is the region covered by both cyan and magenta inks.

If the ink halftone dots are laid out independently, as in Fig. 2, the probability of a light ray to hit the cyan colorant is the probability of hitting the cyan ink multiplied by the probability of not hitting the magenta ink. Similar considerations apply for the other colorants. We obtain

$$\begin{aligned}
 P(\text{Colorant} = \text{purecyan}) &= P(c) \cdot (1 - P(m)) \\
 P(\text{Colorant} = \text{puremagenta}) &= P(m) \cdot (1 - P(c)) \\
 P(\text{Colorant} = \text{blue}) &= P(c) \cdot P(m) \\
 P(\text{Colorant} = \text{white}) &= (1 - P(c)) \cdot (1 - P(m))
 \end{aligned} \tag{1}$$

Since for an incoming photon the probability to hit a given colorant is proportional to the surface coverage of that colorant, we can deduce the surface coverages of the individual colorants

$$\begin{aligned}
a_c &= P(\text{Colorant} = \text{purecyan}) = P(c) \cdot (1 - P(m)) = c \cdot (1 - m) \\
a_m &= P(\text{Colorant} = \text{puremagenta}) = P(m) \cdot (1 - P(c)) = (1 - c) \cdot m \\
a_b &= P(\text{Colorant} = \text{blue}) = P(c) \cdot P(m) = c \cdot m \\
a_w &= P(\text{Colorant} = \text{white}) = (1 - P(c)) \cdot (1 - P(m)) = (1 - c) \cdot (1 - m)
\end{aligned} \tag{2}$$

This calculation of colorant surfaces has been introduced by Demichel [3] and further elaborated by Neugebauer [4].

Along the same reasoning lines, with 3 inks, we obtain Eq. 3 expressing the 8 colorant surface coverages as a function of the cyan (c), magenta (m), and yellow (y) ink surface coverages. The

colorants are white, cyan, magenta, yellow, red (superposed magenta and yellow), green (superposed cyan and yellow), blue (superposed magenta and cyan), and black (superposed cyan, magenta, and yellow) with respective surface coverages a_w , a_c , a_m , a_y , a_r , a_g , a_b , and a_k .

$$\begin{aligned}
\text{white : } a_w &= (1 - c) (1 - m) (1 - y); & \text{cyan : } a_c &= c (1 - m) (1 - y) \\
\text{magenta : } a_m &= (1 - c) m (1 - y); & \text{yellow : } a_y &= (1 - c) (1 - m)y \\
\text{red : } a_r &= (1 - c) m y; & \text{green : } a_g &= c (1 - m)y \\
\text{blue : } a_b &= c m (1 - y); & \text{black : } a_k &= c m y
\end{aligned} \tag{3}$$

With 4 inks we obtain similar expressions (4) for the surface coverages of the 16 colorants, where k is the surface coverage of the black ink:

$$\begin{aligned}
\text{white : } a_w &= (1 - c) (1 - m) (1 - y) (1 - k); & \text{cyan : } a_c &= c (1 - m) (1 - y) (1 - k) \\
\text{magenta : } a_m &= (1 - c) m (1 - y) (1 - k); & \text{yellow : } a_y &= (1 - c) (1 - m)y (1 - k) \\
\text{red : } a_{my} &= (1 - c) m y (1 - k); & \text{green : } a_{cy} &= c (1 - m)y (1 - k) \\
\text{blue : } a_{cm} &= c m (1 - y) (1 - k); & \text{chromatic black : } a_{cmy} &= c m y (1 - k) \\
\text{black : } a_w &= (1 - c) (1 - m) (1 - y) k; & \text{cyanBlack : } a_{ck} &= c (1 - m) (1 - y) k \\
\text{magentaBlack : } a_{mk} &= (1 - c) m (1 - y)k; & \text{yellowBlack : } a_{yk} &= (1 - c) (1 - m)y k \\
\text{redBlack : } a_{myk} &= (1 - c) m y k; & \text{greenBlack : } a_{cyk} &= c (1 - m) y k \\
\text{blueBlack : } a_{cmk} &= c m (1 - y) k; & \text{totalBlack : } a_{cmyk} &= c m y k
\end{aligned} \tag{4}$$

These equations are valid in all cases where the ink halftone dots are laid out independently, e.g.,

in stochastic screening, in error diffusion, and in mutually rotated clustered dot or line screens [5].

References

1. Wyble, D.R., Berns, R.S.: A critical review of spectral models applied to binary color printing. *Color. Res. Appl.* **25**, 4–19 (2000)
2. Hébert, M., Hersch, R.D.: Review of spectral reflectance models for halftone prints: principles, calibration, and prediction accuracy. *Color. Res. Appl.* **40**(4), 383–397 (2015)
3. Demichel, M.E.: *Procédé* **26**(3), 17–21 (1924)
4. Neugebauer, H.E.J.: Die theoretischen Grundlagen des Mehrfarbendrucks. *Z wissenschaftliche Photogr.* **36**, 36–73 (1937)., translated by Wyble, D., Kraushaar, A.: The theoretical basis of multicolor letterpress printing. *Color Res. Appl.* **30**, 323–331 (2005)
5. Amidror, I., Hersch, R.D.: Neugebauer and Demichel: dependence and independence in n-screen superpositions for colour printing. *Color. Res. Appl.* **25**, 267–277 (2000)

Depth in Color

► [Chromostereopsis](#)

Designer's Color Model

► [RYB Color](#)

Deuteranopia

Galina V. Paramei¹ and David L. Bimler²

¹Department of Psychology, Liverpool Hope University, Liverpool, UK

²School of Psychology, Massey University, Palmerston North, New Zealand

Synonyms

[Color blindness](#); [Congenital color vision abnormality](#); [Daltonism](#); [Dichromacy](#); [“Green-blindness”](#); [Red-green deficiency](#); [X-linked inherited color vision deficiency](#)

Definition

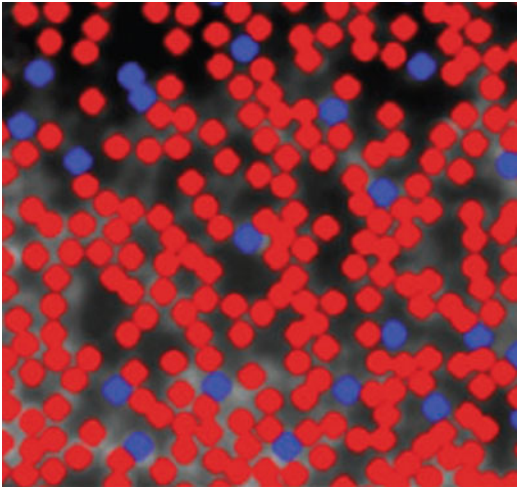
Deuteranopia (from the Greek *deuteros*: second + *an*: not + *opia*: a visual condition) is a congenital form of severe color vision deficiency (dichromacy) affecting the red-green opponent color system. Deuteranopes do not distinguish between colors along a specific direction (deuteranopic confusion lines) in color space and are able to match all colors using two primaries (unlike trichromats who require three primaries). Deuteranopia is linked to the X-chromosome and arises from loss or alteration of the gene encoding the opsin of the middle-wavelength (M-) photopigment, this malstructure resulting in the absence of functioning M-cones in the retina. Deuteranopia follows a recessive pedigree pattern, with incidence of ca. 1% in the Caucasian male population. It is distinct from deuteranomaly in which some degree of discrimination among such colors remains.

Retinal Mosaic

At the level of retinal function, deuteranopia is the absence of M-cones (Fig. 1) (► [Cone Fundamentals](#)), i.e., cone cells expressing a photopigment containing an M-opsin component, tuned for selective sensitivity to medium- to longer-wavelength light such as green and yellow [1].

Genetic Background

Functional loss of the L-cone photoreceptor cells is caused by a mutant cone photopigment gene on the X-chromosome (► [Genetics of Photoreceptors, Genetics and Color Vision Deficiencies, Genes of Cone Photopigments, Genes and Cones](#)). The gene coding for the M-opsin may be missing altogether from a deuteranope's genome; it may be replaced by a second copy of the L-opsin (sensitive to longer wavelengths); or it may be present but damaged so that the protein is not expressed or cannot be transported to the ciliary segment of cone cells expressing that gene [1].



Deutanopia, Fig. 1 Adaptive optics retinal imaging: pseudocolor image of the deutanopic cone mosaic, all cones are of the L- (“red”) or S- (“blue”) type. (Source: Carroll et al. [1]; Fig. 4b, p. 8465. Copyright (2004) National Academy of Sciences, USA)

When present, one or more copies of the M-opsin gene are located on the X-chromosome, preceded by one copy of the L-opsin gene (► [Genetics of Photoreceptors, Genetics and Color Vision Deficiencies, Genes of Cone Photopigments, Genes and Cones](#)). This architecture, and the high similarity of the two genes, facilitates a recombination of DNA during meiotic crossover, leading to deleterious mutations of the M-opsin-encoding gene or its replacement by the L-opsin [2].

With two X-chromosomes, females are far less prone to deutanopia. Notably, mothers and daughters of deuteranopes can be heterozygous for deutanopia (i.e., carriers of the condition) with one normal and one aberrant X-chromosome; the former is active in a portion of their cone cells, with normal and abnormal areas interspersed across the retina. Despite having enough M-cones to sustain normal trichromatic vision, deutanopia carriers manifest subtle departures from it, with reduced subjective prominence of red-green dissimilarity [3, 4], more frequent Ishihara test errors, and wider anomaloscope matching ranges [5] (► [Color Vision Testing](#)).

Confusion Lines

L-signal/M-signal comparison in bipolar ganglion color-comparison cells is not present in deuteranopes (Color Vision, Opponent Theory); hence, there is no difference in appearance of any colors which differ only in the L–M signal. That is, although deuteranopes lack the M- or “green” cones, they are not “green-blind” in a literal sense but rather are red-green-blind. Examples of colors confused by deuteranopes are red, yellow, and green or cyan, gray, and purple. Plotted in a chromaticity diagram (► [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data and Recent Assessments](#)), the confused colors lie on straight “confusion lines” that converge at the deutanope copunctal point [6] (Fig. 2).

Luminous Efficiency and Wavelength Discrimination

The absence of M-cones has little effect on deuteranopes’ sensitivity to colors. Their luminosity function has peak sensitivity around 560 nm, compared to the normal trichromatic peak near 555 nm, slightly redshifted toward the λ_{max} of the L-cone pigment [7, 8].

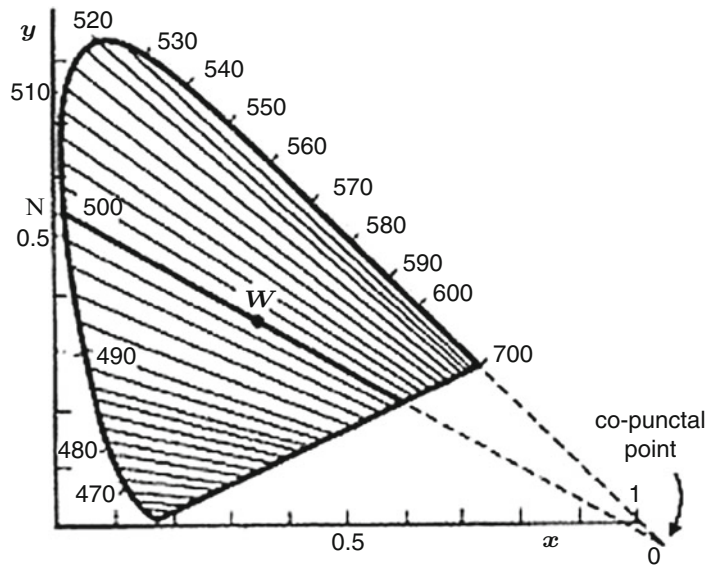
Hue discrimination, as a function of wavelength, is considerably worse for deuteranopes than for normal trichromats. Their lowest discrimination thresholds ($\Delta\lambda$), approaching the trichromatic value, are in the deutanopic “neutral point,” 495–500 nm (see below). $\Delta\lambda$ rises on both sides of this minimum, increasing rapidly on the short-wavelength side ($\lambda < 480$ nm) and on the long-wavelength side ($\lambda > 520$ nm) until wavelengths are not distinguished beyond approximately 530 nm [7].

Color Perception

A deutanope recognizes far fewer distinct colors than a trichromat (► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)). Deutanopic naming of spectral colors reveals a

Deuteranopia,

Fig. 2 Confusion lines indicate colors that cannot be distinguished by a deuteranope. The line going through the white point (W) indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing M-cone fundamental. Drawn in the CIE 1931 chromaticity diagram [6]



“neutral zone” of subjective gray ranging between 495 and 505 nm, with midpoint at ca. 498 nm [9]. For a deuteranope, all equiluminant colors, instead of being arranged in a two-dimensional plane (► [Psychological Color Space and Color Terms](#)), form a single-dimensional line, running between the two extremes of saturated blue and saturated yellow [10]. Deuteranopes distinguish approximately 31 steps along this line, each corresponding to a separate band of interchangeable colors [8]. The usually separate chromatic qualities of “saturation” and “hue” are thus combined along this line. This deuteranopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

Whether the subjective qualities of the two extremes are identical to the normal trichromatic “blue” and “yellow” remains controversial and rests on reports from a handful of individuals with unilateral dichromacy. The blue-yellow convention is used in dichromacy simulation software, which converts color scenes to show how a deuteranope would see these [11].

Despite this model of a two-hue palette for self-luminous colors, the color percepts of deuteranopes remain in question [12]. Their object-color palette was proved to include all six Hering component colors (red, green, blue, yellow, white, and

black) [13]. However, in contrast to normal trichromats, deuteranopes perceive only a weak green component in object colors and do not experience “unique green”; their red and green components are confounded with an achromatic element (► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)).

Diagnosis of Deuteranopia

Deuteranopia can be diagnosed using several tests varying in sensitivity, selectivity, and ease of use [14] (► [Color Vision Testing](#)). Pseudoisochromatic plates and arrangement/panel tests directly address a deuteranope’s confusions between color pairs along the confusion axis. Both types of tests have the advantage of rapid administration and ease of interpretation, making them suitable for field studies or epidemiological applications.

In the pseudoisochromatic tests, Ishihara plates and Hardy-Rand-Rittler test (► [Pseudoisochromatic Plates](#)), each diagnostic plate shows a normal observer design (digit or curved lines) defined by a color difference which disappears for deuteranopes, so that the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction.

In the arrangement tests (Farnsworth-Munsell 100 hue test, Farnsworth D-15 test), the one-dimensionality of deuteranope's color "plane" collapses a color circle into a line (► [Color Circle](#)), causing drastic departures from a normal observer's sequence. Errors, or cap transposition "distances," are summed to quantify the severity of deuteranopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the deuteranopic angle of the confusion axis (► [Color Perception and Environmentally Based Impairments](#)).

The "gold standard" of diagnosis of deuteranopia is the Rayleigh match on a small viewing field (2°) of the Nagel anomaloscope. Conventionally, the levels of 670 nm (red) and 545 nm (yellow) lights are fixed, so that either wavelength at maximum setting stimulates L-cones equally: the so-called deutan mode in which a deuteranope will match *any* red-green ratio to the yellow light by adjusting the luminance [14] (S-cones being unresponsive in this spectral range). Notably, unlike protanopes (► [Protanopia](#)), deuteranopes do not dim the yellow light when adjusting it to the mixed light that is mainly red – the estimate that is diagnostic or differentiating the two red-green deficiency types. Similar to protanopes, when tested with a large viewing field (10°), in setting Rayleigh matches, the majority of deuteranopes perform as anomalous trichromats. Various factors related to retinal eccentricity are presumed, e.g., rod intrusion, S-cone intrusion, spatial variation in the macular pigment, etc. [8].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour Test and Colour Assessment and Diagnosis test (► [Color Vision Testing](#)), interactively quantify any loss of chromatic sensitivity along the deuteranopic confusion line. For a deuteranope, in the frame of the CIE (u' , v') 1976 chromaticity diagram (► [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the deuteranope copunctal point [15, 16].

Incidence of Deuteranopia

Prevalence of inherited deuteranopia varies between human populations [17]. For Caucasian populations, the reported incidence of deuteranopia is 1.27% in males and 0.001% in females [8]. Recent surveys of red-green deficiency in populations of different racial origins found the incidence of deuteranopia in Caucasian males to vary between 0.96% and 1.37%. As is the case of protanopia (► [Protanopia](#)), in indigenous native American, Pacific, Asian, and African populations, the incidence of deuteranopia is markedly lower than in Caucasians, about half this rate, and, too, is observed to rise in coastal areas where migrant Europeans settled in historical times [17]. The increasing prevalence of congenital red-green deficiencies, including deuteranopia, has tentatively been explained by a relaxation in selection pressures as hunting and gathering cultures evolve toward industrialized societies [8]. More recently, however, it is rather attributed to genetic drift and founder events (variable sampling of the gene pool contingent on the population size) [17].

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision Testing](#)
- [Cone Fundamentals](#)
- [Cone Pigments](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)

References

1. Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D.R.: Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8461–8466 (2004)
2. Neitz, M., Neitz, M.: The genetics of normal and defective color vision. *Vis. Res.* **51**, 633–651 (2011)
3. Hood, S.M., Mollon, J.D., Purve, L., Jordan, G.: Color discrimination in carriers of color deficiency. *Vis. Res.* **46**, 2894–2900 (2006)
4. Bimler, D., Kirkland, J.: Colour-space distortion in women who are heterozygous for colour deficiency. *Vis. Res.* **49**, 536–543 (2009)
5. Konstantakopoulou, E., Rodríguez-Carmona, M., Barbur, J.L.: Processing of color signals in female carriers of color vision deficiency. *J. Vis.* **12**(2), art. 11 (2012). <https://doi.org/10.1167/12.2.11>
6. Pitt, F.H.G.: Characteristics of Dichromatic Vision Medical Research Council Special Report Series, No. 200. His Majesty's Stationery Office, London (1935)
7. Pitt, F.H.G.: The nature of normal trichromatic and dichromatic vision. *Proc. R. Soc. Lond. B.* **132**, 101–117 (1944)
8. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Cone genes, cone pigments, color vision and colorblindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, Ch. 1, pp. 3–52. Cambridge University Press, Cambridge (1999)
9. Massof, R.W., Bailey, J.E.: Achromatic points in protanopes and deuteranopes. *Vis. Res.* **16**, 53–58 (1976)
10. Paramei, G.V., Izmailov, C.A., Sokolov, E.N.: Multi-dimensional scaling of large chromatic differences by normal and color-deficient subjects. *Psychol. Sci.* **2**, 244–248 (1991)
11. Brettel, H., Viénot, F., Mollon, J.D.: Computerized simulation of color appearance for dichromats. *J. Opt. Soc. Am. A.* **14**, 2647–2655 (1997)
12. Broackes, J.: What do the color-blind see? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, Ch. 12, pp. 291–405. MIT Press, Cambridge (2010)
13. Logvinenko, A.: On the colours dichromats see. *Col. Res. Appl.* **39**, 112–124 (2014)
14. Dain, S.J.: Clinical colour vision tests. *Clin. Exp. Optom.* **87**, 276–293 (2004)
15. Regan, B., Reffin, J.P., Mollon, J.D.: Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vis. Res.* **34**, 1279–1299 (1994)
16. Barbur, J.L., Rodríguez-Carmona, M., Harlow, J.A.: Establishing the Statistical Limits of “Normal” Chromatic Sensitivity. Publication CIE030, Ottawa (2006)
17. Birch, J.: Worldwide prevalence of red-green color deficiency. *J. Opt. Soc. Am. A.* **29**, 313–320 (2012)

Diagnosis of Defective Color Vision

► [Color Vision Testing](#)

Diagnostic Color

► [Memory Color](#)

Dichromacy

- [Deuteranopia](#)
- [Evolution of Primate Color Vision](#)
- [Protanopia](#)
- [Tritanopia](#)

Dichromat Color Naming

► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)

Diffuse Lighting

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Definition

Lighting in which the light on a point, plane, or object in the lit space is incident with equal luminous intensities from all directions.

Daylighting

The sun, of course, is the primary source of daylight. When there are clouds in the sky, part of the

primary sunlight is absorbed and part reflected from the clouds. The result is a combination of direct sunlight and secondary reflected light from the clouds. In a situation where the sky is completely covered with clouds, the overcast sky acts as a perfect diffuser, resulting in complete diffuse daylighting. The overcast sky lighting condition is used by architects and lighting designers to calculate the lowest level of daylight contribution in buildings. With a partly overcast or complete clear sky, the contribution will be higher. To standardize these calculations, CIE has introduced the so-called CIE overcast sky with a defined luminance distribution. The luminance of the CIE overcast sky is the highest in the zenith, the point straight above the point under consideration, and decreases to one third at the horizon. For simple noncomputerized calculations, sometimes the CIE uniform sky is used in which the luminance distribution of the sky is uniform.

Electric Lighting

Near complete diffuse electric lighting is obtained from indirect lighting in interiors. Here the electric lighting is directed toward the ceiling to create a near uniform ceiling luminance. The light reaches the working plane only after reflection from the ceiling. Since a white ceiling does usually not have a reflectance higher than some 70%, at least 30% of the light is absorbed at the ceiling.

In some interiors, artificial skies are created with luminous ceilings existing of large-sized luminaires with opal type of louvers. The lighting of these artificial skies is usually diffuse although, with LED or OLED solutions, nonuniform skies and consequently non-diffuse lighting can be created.

Conventional indoor lighting luminaires for general lighting purposes, in fact, have a diffuse component with, under elevation angles of some 60–70°, a directional component. Dependent on the actual light distribution of the luminaire, the lighting is more diffuse or more directional.

Consequences of Diffuse Lighting

The consequences of complete diffuse lighting are best explained by considering the condition of a clouded sky. The diffuse lighting of it is not appreciated very much: it gives a dull, monotonous impression in which shadows, depth, shape, and texture are poor or completely absent. This has negative consequences for both visual pleasantness and visual performance. An illustration of the latter is an entirely cloudy sky on a ski slope: irregularities and bumps in the snow cannot be seen because of the absence of shadows. Indeed it is generally considered dangerous to make a somewhat difficult descend under these viewing conditions.

Diffuse electric lighting has the same negative consequences for both pleasantness and visual performance. The effect of lack of shadows, here, is often referred to as bad modelling.

Good interior lighting consists of both a diffuse and directional component. This again is analogue to a relatively clear sky outdoors with a strong directional component from the sun with some diffuse light from a couple of white clouds.

Cross-References

- [Architectural Lighting](#)
- [Daylighting](#)
- [Interior Lighting](#)

1,2-Dihydroxy-9,10-anthracenedione

- [Alizarin](#)

1,2-Dihydroxyanthraquinone

- [Alizarin](#)

Direct to Garment (DTG)

► [Coloration, Ink-Jet Printing](#)

Displays

► [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

Drop-on-Demand (DOD)

► [Coloration, Ink-Jet Printing](#)

Dry or Mechanical Finishing

► [Finish, Textile](#)

Dye

Jae Hong Choi
Department of Textile System Engineering,
Kyungpook National University, Daegu, South
Korea

Synonyms

[Colorant](#)

Definition

The term *dye* refers to compounds which can impart color to a substrate when applied in solution from either aqueous or organic solvents. The substrates include textiles, plastics, polymers, etc. Both the applications to materials and color constitutions of dyes involve essentially chemical principles. Dyes and pigments are both commonly

applied as a coloring material, but they are distinguished on the basis of their solubility either in water or organic solvents; therefore dyes are soluble, whereas pigments are insoluble.

History of the Development of Synthetic Dyes

Until the middle of the nineteenth century, the dyeing of textiles and leather was carried out using dyes derived from animal and vegetable sources. Of such natural dyes, indigo is perhaps the most well-known and most important material [1]. Alizarin is the most important species of anthraquinone-based red dyes obtained from the roots of plants of the *Rubiaceae*. A highly prized plant was madder which produced dyeings on cotton known as Turkey Red by means of Ca-Al mordant. Also of importance in early times was the use of the red dye, Kermes, obtained from insects [1, 2]. In 1856, WH Perkin synthesized a brilliant violet dye, Aniline Purple (which later became known as Mauveine), by the oxidation of a mixture of aniline bases [3–5]. Perkin's (accidental) discovery of Mauveine and his development of methods to apply the dye to silk and cotton stimulated enormous commercial interest in dye synthesis. Further discoveries quickly followed, such as that of the diazo reaction by Griess in 1859 and the introduction of the first azo dye, Bismarck Brown, by Martius in 1863 [6] and the structural elucidation and synthesis of Alizarin that occurred in 1868 and Indigo by von Baeyer in 1883 and Heumann in 1890; shortly afterwards the first anthraquinoid vat dye indanthrone was prepared by Bohn in 1901 [6, 7]. As a result of intensive attempts to devise new dyes for application to the fiber cellulose acetate, the water-soluble *Ionamin* range of dyes was introduced in 1922 [8] which were the precursors of the current wide range of sparingly soluble disperse dyes. The first reactive dyes for cotton were introduced by ICI in 1956; such dyes are able to attach covalently to cellulosic fiber, thereby offering dyeings of excellent fastness to washing and wet treatments [9]. From the 1990s, non-textile applications for dyes, such as

information technology, reprographics, optoelectronics, and solar cells, have contributed to the diversity of modern synthetic dyes [1, 10].

Principles of Color Constitution in Dyes

When ultraviolet or visible radiation is absorbed, an n or π electron is promoted and occupies an empty π^* or σ^* antibonding orbital, the wavelength, λ , of the photon being equal to the energy difference, ΔE , between the two levels, ground state and excited state, involved:

$$\lambda = hc/\Delta E$$

where h is Planck's constant and c the velocity of light.

When the energy difference is equal to the energy of visible light, the compound has an absorption band in the visible region (400–700 nm) and exhibits the color complementary to that of the absorbed light [7].

Early Theories Using Qualitative Method

The elucidation of the relationship between the color of dyes and their chemical constitution has attracted much attention. In 1876, Witt suggested that color arises from the combination of unsaturated groups, such as $-\text{N}=\text{N}-$, $-\text{C}=\text{O}$, and NO_2 , with basic groups, such as $-\text{NH}_2$, $-\text{OH}$, and $-\text{NR}_2$, within a dye molecule. Witt defined the former groups as *chromophores* and the latter as *auxochromes*. The terminology of *bathochrome* and *hypsochrome* was also coined at the same time [6, 11]. Hewitt defined the role of conjugation in a dye molecule as “the longer the conjugated chain, the more bathochromic the dye” and pointed out the importance of a conjugated chain for the production of color. With the discovery of the electron in 1897 [12], attention was then given to the behavior of electrons within dye structures as providing the origin of color.

Modern Theories

All modern theories involve two distinct approaches, namely, the Valence Bond (VB) and Molecular Orbital (MO) theories, which are developments from quantum mechanics. VB

theory is principally based on bonding valence electron pairs localized between specific atoms in a dye molecule [12], whereas MO theory has played a key role in the development of modern color theories for solving the intricate relationships between color and chemical constitution. A satisfactory method to calculate the transition energy and intensity of visual color was developed in terms of the Linear Combination of Atomic Orbitals (LCAO) method and the Hückel Molecular Orbital (HMO) method, which were originally developed for unsaturated hydrocarbons. Hückel assumed that the transition state would correspond to the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

Classification and Nature of the Chromogens

Colored organic molecules can be rationalized by classification into four groups [13]:

1. $n \rightarrow \pi^*$ chromogens
2. Donor–acceptor chromogens
3. Acyclic and cyclic chromogens
4. Cyanine-type chromogens

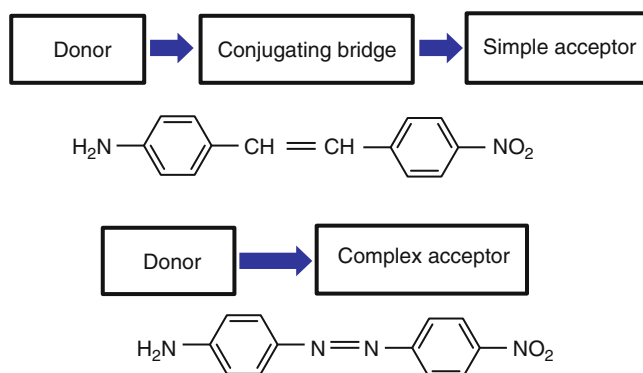
Thus, dyes which exhibit electronic transitions from a lone pair, nonbonding atomic orbital of the molecule belong to the $n \rightarrow \pi^*$ group of chromogens. However, most dyes are regarded as donor–acceptor chromogens, which, characteristically, possess electron donor and electron acceptor groups that are linked by a conjugated π -electron system (the conjugating bridge). Depending on the role of the accepting group, donor–acceptor chromogens can be subdivided into simple acceptors or complex acceptors, as represented in Fig. 1 [13].

For azo dyes, in general, the strength of the electron acceptor correlates closely with the Hammett σ constant of the substituent. As the value of the σ constant of a substituent increases, the electron acceptor group exerts a more powerful effect and produces a larger bathochromic shift (absorption maximum moves to longer wavelength).

Effect of Substituents on Color in Azo Dyes

In simple azo dyes, a weak long-wavelength absorption is observed due to the overlap of the

Dye, Fig. 1 Schematic representation of simple acceptor and complex acceptor



D

atomic orbitals of the two nitrogen atoms, giving rise to an $n \rightarrow \pi^*$ transition. The intensities of $n \rightarrow \pi^*$ transitions vary depending on whether the transition is symmetrically allowed or forbidden [14]. A weak absorption band, involving $n \rightarrow \pi^*$ transition, tends to be overshadowed by the more intense $\pi \rightarrow \pi^*$ absorption band. If an electron-donating group, such as an dialkylamino group, is introduced into the cyclic unsaturated residue (**D**) of an azo dye, of general formula $A - N = N - D$, an intense new absorption band arises as a result of a $\pi \rightarrow \pi^*$ transition, which is referred to as a charge-transfer (CT) transition. The migration of a π electron delocalized in the nitrogen orbital of the amino group towards the azo dye is responsible for such charge-transfer transitions [13]. A further bathochromic shift can be exerted by the additional incorporation of electron-withdrawing groups, such as NO_2 , CN, and halogen, into the **A** ring. The greatest effect in terms of longer wavelength is achieved by placing the substituents at positions *ortho*- or *para*- to the azo group, which maximizes conjugation. Similarly, electron-donating groups situated *ortho* to the azo linkage in the **D** ring normally give rise to the most effective bathochromic shifts [13, 15].

Principles of Photofading in the Dye Molecule

Jablonski Diagram

Theoretically, once a dye molecule absorbs a photon from incident light, an electron will be excited to different energy levels. As depicted in Fig. 2,

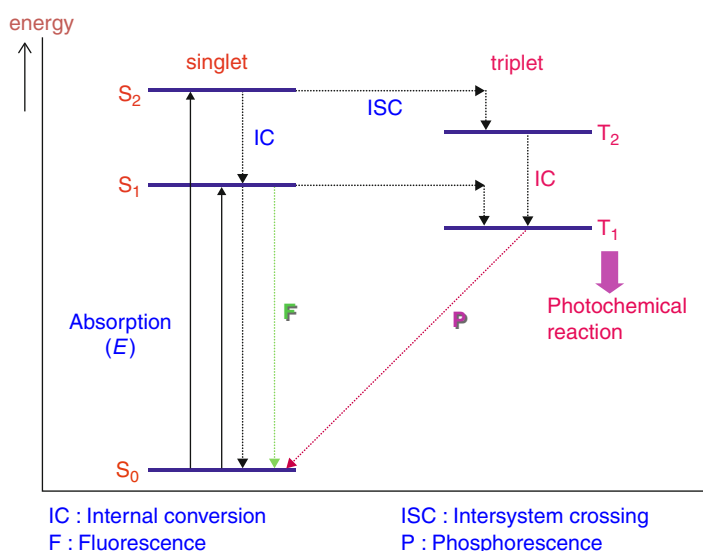
the excited molecule tends to lose its energy and returns to the ground state, this process involving the emission of fluorescence or phosphorescence, or can undergo a chemical reaction, leading to decomposition. This type of reaction is referred to as a *photochemical reaction* or photofading. Not surprisingly, both the first singlet excited state and the first triplet excited state are the crucial energy levels for this reaction. The longer lifetime of the triplet excited state may be linked with a higher likelihood of photodegradation of the dye molecule [6, 13].

Internal conversion of the excited state is expected to be the fastest step for dyes and UV stabilizers that are employed in polymers, whereas fluorescence should be both dominant and fastest for optical brightening agents and laser dyes. Intersystem crossing should be fastest for dyes used as sensitizers [6].

Types of Photoreaction

Photofading may occur in different ways depending on the nature of the substrates, wavelength distribution of the incident light, the presence of moisture, and air and dye structure. In general, two principal types of photochemical reaction are recognized: photooxidation and photoreduction. The influence of substituents on degradation rate has been rationalized by plotting photofading rate against Hammett σ constant for aromatic azo dyes containing various substituents. Thus, if the fading rate increases as electron-withdrawing power increases (higher σ value), a reductive mechanism is favored (positive), whereas a decreased fading rate is observed when oxidative degradation dominates

Dye, Fig. 2 Energy-level scheme of a photoexcited dye molecule (Jablonski diagram)



(negative) [11, 16]. As far as photooxidation is concerned, singlet oxygen plays a very important role.

Solvatochromism of Dyes

Solvatochromism can best be described as the effect of a solvent on the color of a dye. Solvent–solute interactions have various influences on the absorption spectrum, such as λ_{max} value and the intensity of a band; the bandwidth may also vary.

Solvatochromism Types

Generally, in many dye molecules, the ground state is less polar than the excited state so that a polar solvent will tend to stabilize the excited state more than the ground state, leading to a bathochromic shift in the absorption maximum; this effect is termed positive solvatochromism. In contrast to positive solvatochromism, negative solvatochromism is relatively rare and is most often observed in dyes that contain a delocalized charge [6]. The interaction of a solvent with a dye molecule is greatest in polar solvents, for example, ethanol, which possess a strong permanent dipole, and is most pronounced with a solute molecule which contains a permanent dipole. Intermolecular hydrogen bonding can also give rise to shifts in the absorption spectrum,

particularly in case of $n \rightarrow \pi^*$ bands. For heteroatoms, such as nitrogen, sulfur, and oxygen, the nonbonding lone pair electrons can form intermolecular hydrogen bonds with suitable protic polar solvents, such as ethanol or water. Accordingly, the excitation energy will be raised by the hydrogen bond, which stabilizes the n -electrons resulting in a hypsochromic shift in the absorption spectrum [13, 17].

Halochromism of Dyes

This term is defined as the color change of a substance accompanying a change in the pH of the solution [13]. Although many types of aminoazo dyes show this effect, it is observed only when the amino group is situated *para* to the azo linkage.

Halochromism Types

In contrast to hydroxy azo dyes, in acidic solution, many dyes derived from 4-aminoazobenzene are susceptible to protonation either at the β -nitrogen atom of the azo linkage, to give the azonium cation, or at the amino nitrogen atom to produce an ammonium ion. Bathochromic shifts combined with increased color intensity are often observed for the azonium ion. This shift is termed as a positive halochromism, whereas the opposite

3. Rys, P., Zollinger, H.: *Fundamentals of the Chemistry and Application of Dyes*. Wiley-Interscience, London (1990)
4. Peter, J.T., Anthony, S.T.: A history of the international dyestuff industry. *Am. Dyestuff Rep.* **81**(11), 59–100 (1992)
5. Welham, R.D.: The early history of the synthetic dye industry: part 2 the industrial history (1856–1900). *J. Soc. Dye Col.* **79**(3), 146–152 (1963)
6. Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments*. VCH/VCH-VCH, Zürich (2003)
7. McLaren, K.: *The Colour Science of Dyes and Pigments*. Adam Hilger Ltd, Bristol (1986)
8. Green, A.G., Saunders, K.H.: The ionamines: a new class of dyestuffs for acetate silk. *J. Soc. Dye Col.* **39**(1), 10 (1923)
9. Shore, J. (ed.): *Colorants and Auxiliaries*, vol. 1. SDC, Bradford (1990)
10. Kim, S.H. (ed.): *Functional Dyes*. Elsevier, Amsterdam (2006)
11. Griffiths, J. (ed.): *Developments in the Chemistry and Technology of Organic Dyes*. Blackwell Scientific Publications, Oxford (1984)
12. Gordon, P.F., Gregory, P.: *Organic Chemistry in Colour*. Springer, Berlin (1987)
13. Griffiths, J.: *Colour and Constitution of Organic Molecules*. Academic Press, London (1976)
14. Fabian, J., Hartmann, H.: *Light Absorption of Organic Colorants*. Springer, Berlin (1980)
15. Griffiths, J., Roozpeikar, B.: Synthesis and electronic absorption spectra of dicyano-derivatives of 4-diethylaminoazobenzene. *J. Chem. Soc. Perkin Trans. I*, 42–45 (1976)
16. Giles, C.H., Hojiwala, B.J., Shah, C.D.: Quantum efficiency measurements of fading of some disperse dyes in nylon and polyester films and in solution. *J. Soc. Dye Col.* **88**(11), 403–407 (1972)
17. Christian, R., Thomas, W.: *Solvents and Solvent Effects in Organic Chemistry*, 4th edn. Wiley-VCH, Weinheim (2010)
18. Griffiths, J., Hill, J., Fishwick, B.: The application of PPP-MO theory to the halochromism of 4-aminoazobenzene dyes. *Dyes Pigments*. **15**(4), 307–317 (1991)
19. Bello, K.A., Griffiths, J.: Violet to cyan azo dyes derived from 4-amino-3-nitrobenzaldehyde as diazo component. *Dyes Pigments*. **11**(1), 65–76 (1989)
20. Hallas, G.: The effect of terminal groups in 4-aminoazobenzene and disperse dyes. *J. Soc. Dye Col.* **95**(8), 285 (1979)
21. Kuehni, R.G.: *Color: an Introduction to Practice and Principles*. Wiley-Interscience, Hoboken (2005)

Dye Functional

► [Colorant, Nonlinear Optical](#)

Dye, DSSC

Mira Abed and Ahmed El-Shafei

Department of Textile Engineering, Chemistry & Science, North Carolina State University, Raleigh, NC, USA

Synonyms

[Dye-sensitized solar cells](#)

Definition

Dye-sensitized solar cells (DSSCs) are thin-film photovoltaic devices that use dyes covalently bonded to a semiconductor surface to harvest energy (photons) from the sun and convert it into electrical energy. In DSSC fabrication, a dye-containing thin semiconductor film, typically titanium dioxide, is applied to a conductive glass substrate. The resulting photoanode is attached to a counter electrode, typically a platinum-coated conductive glass. A liquid electrolyte is injected between the two electrodes to complete the circuit and ensure dye regeneration. The dyes used, also known as photosensitizers, are often ruthenium-complexed compounds owing to their exceptional photophysical and photochemical properties and long lived excited states, but other classes used include metal-free dyes and various other high efficiency metal-complexes such as Zn-porphyrins (A3B type). Molecular engineering of photosensitizers for DSSCs has been a significant research focus in the last three decades, as the photosensitizer is responsible for the quantity and energy of harvested photons.

Introduction

Dye-sensitized solar cells (DSSCs) emerged from a research effort aimed at finding alternatives to silicon solar cells, which require high fabrication temperatures. While DSSCs do not perform as efficiently as their silicon counterparts under direct lighting conditions, they perform just as

well or better under all other conditions – in particular under diffuse light [1]. Additionally, they are less expensive and easier to fabricate than silicon solar panels, making the cost per Watt advantageous. Because DSSCs do not require high temperatures to fabricate, they can be made on flexible substrates such as plastics to furnish devices that are flexible and can be also integrated into fabrics [1, 2].

Structure of Dye-Sensitized Solar Cells

The structure of a dye-sensitized solar cell is shown in Fig. 1. The DSSC is constructed by sandwiching an electrolyte between a photoanode and a counter-electrode. Both electrodes are fabricated on conductive glass, which is coated with fluorine-doped tin oxide (FTO) to make it conductive.

Photoanode

The photoanode comprises a metal oxide, typically nanocrystalline TiO_2 , with an adsorbed dye for harvesting solar energy. The excited state of the dye should be at a slightly higher energy level than that of the conduction band edge of the TiO_2 for thermodynamically favorable electron injection from the first excited state singlet and/or triplet of the dye into the conduction band edge of semiconductor.

For fabrication of the photoanode, a conductive FTO-glass is printed with porous TiO_2 . For optimized DSSC efficiency, TiO_2 is deposited in

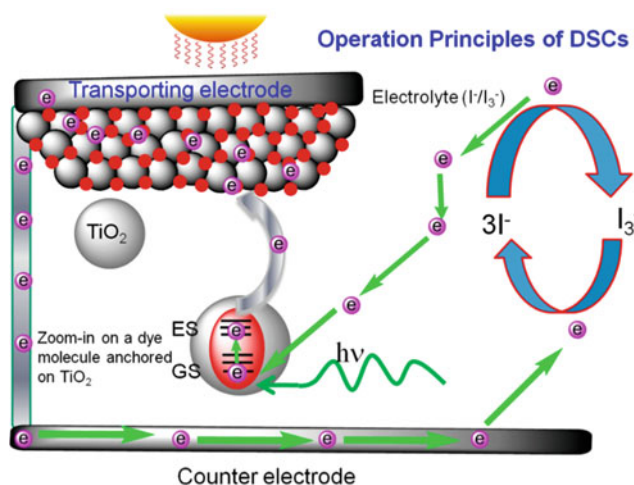
multiple layers, the first of which is called the blocking layer because it minimizes electron transfer between the electrolyte and the conductive FTO on the glass surface [1]. Once the $\sim 50\text{ nm}$ thick blocking layer of TiO_2 nanoparticles has been annealed, one or two mesoporous layers are printed on top. This typically consists of a $\sim 10\text{ }\mu\text{m}$ thick light absorption layer with particle size $\sim 20\text{ nm}$ and a $\sim 3\text{ }\mu\text{m}$ light scattering layer with particle size $\sim 400\text{ nm}$. These two layers allow large surface areas for increased dye adsorption and the pores allow electrolyte diffusion. Finally, a $\sim 1\text{ nm}$ thick pure TiO_2 shell is deposited via aqueous TiCl_4 dipping. This covers any impurities that may be on the mesoporous surface and increases surface roughness (and therefore dye adsorption), lowering the Fermi level, and hence the energy level of the conduction band edge of TiO_2 . The lower (more negative) free energy of the conduction band increases the driving force of electron injection from the excited state of the dye and leads to longer electron lifetimes [1].

Counter Electrode

The counter electrode functions as the cathode of the DSSC and is responsible for reducing I_3^- to I^- at the interface. The counter electrode can be fabricated by printing a layer of catalytic platinum paste on FTO-glass and annealing. These pastes are typically comprised of platinum chloride compounds which thermally decompose to produce Pt nanoclusters [1]. In addition to the cost and energy

Dye, DSSC, Fig. 1

Architecture of a standard dye-sensitized solar cell illustrating light absorption, dye excitation, electrodes, electron injection from the dye into TiO_2 , and role of electrolyte



benefits of using transparent catalytic Pt films, there are also stability benefits, as electro-deposited or vapor-deposited Pt films tend to dissolve in I^-/I_3^- [1]. As an added advantage, the DSSC can be illuminated from either side, increasing the incident light in applications such as windows. Furthermore, the thickness of the active TiO_2 layer is responsible for the cell transparency, which creates a tradeoff between the cell's efficiency and transparency.

Electrolyte

The most common electrolyte system used in DSSCs is an I^-/I_3^- redox couple in an organic solvent. The electrolyte system is responsible for accepting incoming electrons from the platinum counter electrode and for regenerating the dye via reduction. For unsealed devices, the electrolyte is allowed to diffuse between the two layers of glass that make up the “sandwiched” DSSC construction. For sealed devices, a hole is drilled in the counter electrode before the photoanode and counter electrode are sealed together in the sandwich construction using an ionomeric film. The electrolyte solution is then injected into the cavity between the two active layers via the hole in the counter electrode [3].

While the electrolyte is intended to play the beneficial role of dye regeneration, it can also have detrimental effects. Both the I^-/I_3^- and the organic solvent components are volatile, and as they evaporate the effectiveness of the device is diminished until the solution is replenished [4]. Especially problematic in copper-complexed dyes is the potential for a coordinating solvent to bind to the copper center and prevent dye regeneration, resulting in poor stability and efficiency [4]. Additionally, I^-/I_3^- itself absorbs light and can compete with the dye for light absorption. For these reasons, the search for alternatives to this electrolyte system has been a significant area of research [4, 5]. Recently, researchers have investigated cobalt- and copper-based systems. In every case, it is important to choose an electrolyte that is compatible with the chosen dye. For example, a copper-complexed dye and a copper-based electrolyte would not be compatible if ligand exchange occurs, resulting in the

degradation of both the photosensitizer and the electrolyte. The I^-/I_3^- electrolyte system has been optimized for use with ruthenium-complexed dyes in DSSCs.

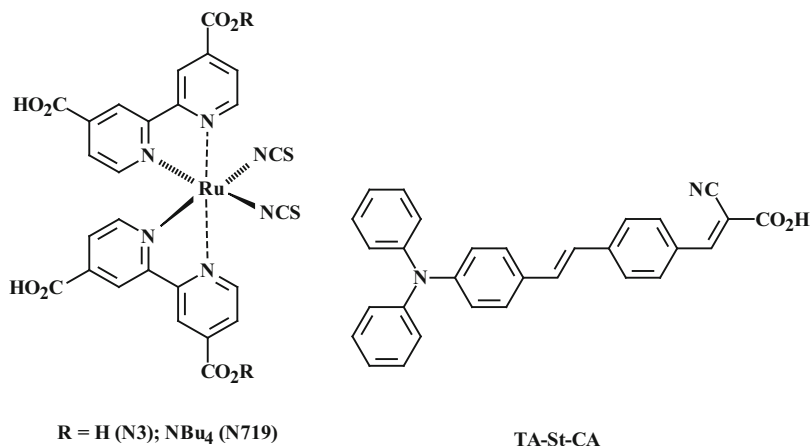
Photosensitizer

The photosensitizer is responsible for light absorption and the resultant generation of excited electrons. When the TiO_2 layer is completed, the TiO_2 -coated glass is soaked in a dye solution for 20–24 h at approximately 25 °C. This dye solution sometimes contains a co-adsorbent to prevent dye aggregation, which can inhibit electron injection. The photoanode is then removed from the dye solution, rinsed, and dried. The dipping process can be more complicated for certain copper-complexed dyes, as the dyes themselves are assembled in-situ and require a three-step dipping process [6]. The photosensitizer is the component of the cell that absorbs light, which excites electrons to an excited state in the process. The mechanism by which the excited dye relaxes to the ground state is of great importance for DSSCs, because many of the available relaxation mechanisms preclude injection of electrons into the metal oxide. Examples of dyes that have been developed for commercial DSSC devices are shown in Fig. 2. These dyes have a photon to current conversion efficiency (η) of 10% (N3), 11.18% (N719), 9.1% (TA-St-CA) [7, 8].

Ru (II) complexes exhibit strong spin-orbit coupling [9], and their resulting long-lived excited states have proven advantageous for photocurrent generation in DSSCs [9–13]. Since the first report of an Ru-based DSSC by Michael Grätzel in 1991 [13], a significant research effort has focused on structural modifications to Ru(II) polypyridyl complexes for optimized function in DSSCs [14, 15]. In addition, several other broad classes of dyes have emerged in recent years. Organic (metal-free) dyes have been a popular focus, with the benefit of eliminating the need for rare and expensive Ru while achieving high molar extinction coefficients and conversion efficiencies [1, 16]. Porphyrin dyes recently made a dramatic entrance to the field with 11% efficiency in 2010 [17] and have since reached 13% [18]. There has also been some work on ruthenium-free metal-

Dye, DSSC, Fig. 2

Examples of
photosensitizers used in
commercial DSSCs



based dyes, including those based on osmium, copper, iron, platinum, and rhenium [1].

There are several basic structural requirements of dyes for metal-based DSSCs. The first and most important are anchoring groups for TiO₂ adsorption and electron injection. In Ru(II) complexes for DSSCs, these are typically acidic functional groups such as carboxylic or phosphonic acid groups, but can include sulfonic acid, cyanoacrylic acid, and others [19]. Some devices have been reported with no such anchoring groups, but they exhibit very poor electron injection and stability [20]. While not strictly required for a metal-based sensitizer, strong electron-donating substituents are nearly ubiquitously employed on ancillary ligands to broaden light absorption via destabilization of the HOMO [21].

Working Function and Principles

A dye-sensitized solar cell operates as shown in Fig. 3. First, incident light excites an electron within the photosensitizer (step 1 in Fig. 3). If the thermodynamic driving force ($-\Delta G$) is sufficient, the electron is injected from the excited state of the dye into the conduction band edge of the TiO₂ layer (step 2). If the electron successfully avoids recombination (step 5), it moves into the external circuit. Once the electron reaches the platinum counter electrode, it then enters the cell via reduction of the I^-/I_3^- couple at the Pt/electrolyte interface (step 4). The electrolyte subsequently regenerates the original dye via reduction (step 3), and the cycle

is complete [15]. The red arrows in Fig. 3 represent processes that decrease the efficiency of the DSSC. Step 5 represents an electron that is injected into the conduction band of TiO₂ but subsequently recombines with the dye to regenerate the ground state. Step 6 represents reduction of the redox mediator by electrons in the conduction band of TiO₂ [14].

Thermodynamic considerations are critical when selecting the components of a DSSC. The excited state oxidation potential (ESOP) of the dye must be at least 150–200 mV greater than the conduction band of the metal oxide for favorable electron injection. Additionally, the HOMO of the dye must be at least 200–300 mV lower than that of the redox potential of the electrolyte for favorable dye regeneration via reduction [14, 22].

Photovoltaic Parameters

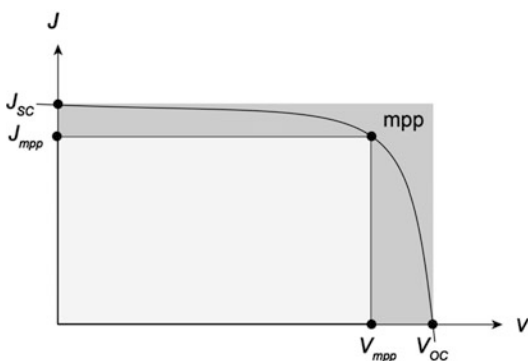
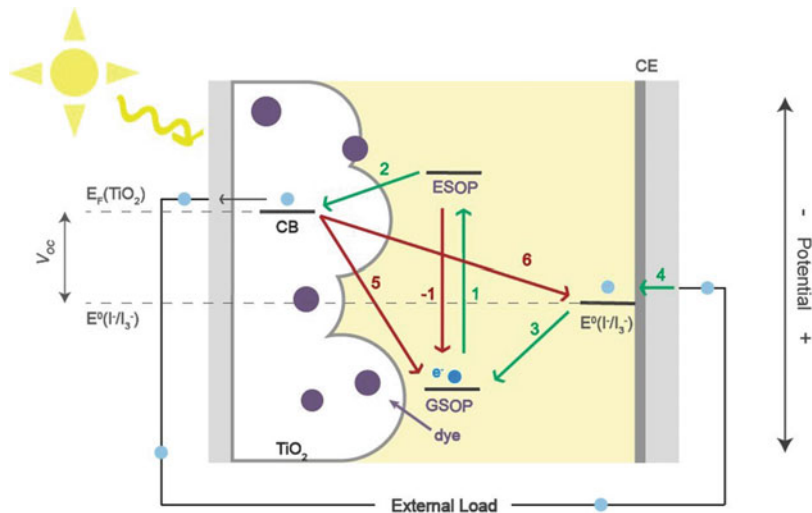
The most commonly cited photovoltaic parameter is the power conversion efficiency (PCE or η). It is defined as the maximum output power as a percent of the input power, as shown in Eq. 1 [1, 23]:

$$\eta = \frac{P_{\max}}{P_{\text{in}}} \times 100\% \quad (1)$$

The power conversion efficiency is obtained from current density-voltage measurements. The cell is irradiated with a one Sun solar simulator and current density is measured at each applied voltage, generating current density-voltage curves such as that shown in Fig. 4 [23].

Dye, DSSC, Fig. 3

Working principle of DSSCs



Dye, DSSC, Fig. 4 Illuminated current density-voltage curve of an ideal solar cell

J_{sc} is the short circuit photocurrent density, where $V = 0$ and each excited electron is injected virtually instantaneously after generation. V_{oc} is the open circuit photovoltage, where $J = 0$ and every excited electron recombines rather than generating current. At some intermediate point along the $J - V$ curve is the maximum power point (mpp), where $J = J_{mpp}$, $V = V_{mpp}$, and $P = P_{max}$. As we can see from Fig. 4:

$$P_{max} = J_{mpp} V_{mpp} \quad (2)$$

$$\frac{P_{max}}{J_{sc} V_{oc}} = \frac{J_{mpp} V_{mpp}}{J_{sc} V_{oc}} = FF \quad (3)$$

The ratio in Eq. 3 is defined as the fill factor (FF), and allow us to express the efficiency in terms of J_{sc} and V_{oc} :

$$\eta = \frac{J_{sc} V_{oc} FF}{P_{in}} \quad (4)$$

The fill factor represents the fraction of maximum theoretical power the solar cell can achieve, and reflects the overall quality of the device.

Final Considerations

A significant portion of DSSC research has involved the design of new sensitizers, with the goal of broadening solar energy harvesting, improving electron injection dynamics, increasing excited state lifetimes, and improving stability, all with the goal of increasing long-term efficiency of DSSCs. Four main classes of dyes suitable for DSSCs have emerged: metal-based coordinate complexes (mostly ruthenium-based), organic dyes (metal-free), Zn-porphyrin dyes, and natural dyes.

Design Features of Photosensitizers

Dyes for DSSCs must be structurally suited to the application by satisfying the following requirements: (1) the dye must exhibit affinity for the metal oxide, which is accomplished through the incorporation of anchoring groups; (2) the HOMO (GSOP) and ESOP must align with the conduction band edge of TiO_2 and the redox potential of I^-/I_3^- , respectively, for thermodynamically favorable electron injection and dye regeneration;

(3) the LUMO should be centered on the anchoring ligand for kinetically favorable electron injection into the semiconductor; (4) the dye must be not only photostable, it must also be chemically stable over at least 10^8 oxidation-reduction cycles. This corresponds to an approximated 20 years of continuous irradiation [14].

Different dye classes employ different structural features to achieve these goals. Strong electron donors such as carbazole and triphenylamine units are prevalent, especially in organic dyes and transition metal coordinate complexes [1, 17]. Donor-free thiophene derivatives have also become popular substituents because they function as hole-transporting domains and increase conductivity within the device [14].

Ruthenium complexes are by far the most popular photosensitizers for a variety of applications, and DSSCs are no exception. Ruthenium polypyridyl complexes have high molar absorptivities, strong MLCT absorption bands in the visible range, and long excited state lifetimes due to an easily accessible triplet state, making them the long-standing prototype [15].

Cross-References

- [Chromophore](#)
- [Coloration, Mordants](#)
- [Colour Index](#)
- [Dye](#)

References

1. Hagfeldt, A., Boschloo, G., Sun, L., Kloo, L., Pettersson, H.: Dye-sensitized solar cells. *Chem. Rev.* **110**, 6595–6663 (2010)
2. Liu, J., Li, Y., Arumugam, S., Beeby, S.: Screen printed dye-sensitized solar cells (DSSCs) on woven polyester cotton fabric for wearable energy harvesting applications. *Mater. Today Proc.* **5**, 13753–13758 (2018)
3. Brauchli, S.Y., Constable, E.C., Housecroft, C.E.: Concentration effects on the performance of bis(diimine) copper(I) dyes in dye-sensitized solar cells. *Dyes Pigments*. **113**, 447–450 (2015)
4. Ashbrook, L.N., Elliott, C.M.: Dye-sensitized solar cell studies of a donor-appended bis(2,9-dimethyl-1,10-phenanthroline) Cu(I) dye paired with a cobalt-based mediator. *J. Phys. Chem. C*. **117**, 3853–3864 (2013)
5. Karthikeyan, B.C.S., Wietasch, H., Thelakkat, M.: Highly efficient solid-state dye-sensitized TiO₂ solar cells using donor-antenna dyes capable of multistep charge-transfer cascades. *Adv. Mater.* **19**, 1091–1095 (2007)
6. Schönhofer, E., et al.: “Surfaces-as-ligands, surfaces-as-complexes” strategies for copper(I) dye-sensitized solar cells. *Dyes Pigments*. **115**, 154–165 (2015)
7. Nazeeruddin, M.K., et al.: Combined experimental and DFT-TDDFT computational study of photoelectrochemical cell ruthenium sensitizers. *J. Am. Chem. Soc.* **127**, 16835–16847 (2005)
8. Hwang, S., et al.: A highly efficient organic sensitizer for dye-sensitized solar cells. *Chem. Commun.* **21**, 4887–4889 (2007)
9. Campagna, S., Puntoriero, F., Nastasi, F., Bergamini, G., Balzani, V.: In: Balzani, V., Campagna, S. (eds.) *Photochemistry and Photophysics of Coordination Compounds I*, pp. 117–214. Springer, Berlin/Heidelberg (2007)
10. Benkö, G., Kallioinen, J., Korppi-Tommola, J.E.I., Yartsev, A.P., Sundström, V.: Photoinduced ultrafast dye-to-semiconductor electron injection from non-thermalized and thermalized donor states. *J. Am. Chem. Soc.* **124**, 489–493 (2002)
11. Benkö, G., et al.: Interligand electron transfer determines triplet excited state electron injection in RuN3-sensitized TiO₂ films. *J. Phys. Chem. B*. **108**, 2862–2867 (2004)
12. Cheema, H., et al.: Influence of mono versus bis-electron-donor ancillary ligands in heteroleptic Ru(II) bipyridyl complexes on electron injection from the first excited singlet and triplet states in dye-sensitized solar cells. *J. Mater. Chem. A*. **2**, 14228 (2014)
13. O'Regan, B., Grätzel, M.: A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO₂ films. *Nature*. **353**, 737–740 (1991)
14. Pashaei, B., Shahroosvand, H., Graetzel, M., Nazeeruddin, M.K.: Influence of ancillary ligands in dye-sensitized solar cells. *Chem. Rev.* **116**, 9485–9564 (2016)
15. Grätzel, M.: Recent advances in sensitized mesoscopic solar cells. *Acc. Chem. Res.* **42**, 1788–1798 (2009)
16. Lee, C.P., et al.: Recent progress in organic sensitizers for dye-sensitized solar cells. *RSC Adv.* **5**, 23810–23825 (2015)
17. Bessho, T., Zakeeruddin, S.M., Yeh, C., Diau, E.W., Grätzel, M.: Highly efficient mesoscopic dye-sensitized solar cells based on donor-acceptor-substituted porphyrins. *Angew. Chem. Int. Ed.* **49**, 6646–6649 (2010)
18. Mathew, S., et al.: Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nat. Chem.* **6**, 242–247 (2014)
19. Zhang, L., Cole, J.M.: Anchoring groups for dye-sensitized solar cells. *ACS Appl. Mater. Interfaces*. **7**, 3427–3455 (2015)
20. Sakaki, S., Kuroki, T., Hamada, T.: Synthesis of a new copper(I) complex, [Cu(tmcbpy)₂]⁺(tmcbpy[−] =

- 4,4',6,6'-tetramethyl-2,2'-bipyridine-5,5'-dicarboxylic acid), and its application to solar cells. *J. Chem. Soc. Dalt. Trans.* **2002**, 840–842 (2002)
21. Nazeeruddin, M.K., et al.: DFT-INDO/S modeling of new high molar extinction coefficient charge-transfer sensitizers for solar cell applications. *Inorg. Chem.* **45**, 787–797 (2006)
 22. Cheema, H., Islam, A., Han, L., El-Shafei, A.: Monodentate pyrazole as a replacement of labile NCS for Ru (II) photosensitizers: minimum electron injection free energy for dye-sensitized solar cells. *Dyes Pigments.* **120**, 93–98 (2015)
 23. Bauer, G.H.: *Photovoltaic Solar Energy Conversion*, p. 901. Springer, Berlin/Heidelberg (2015)

Dye, Functional

Sung-Hoon Kim and Eun-Mi Lee
Department of Textile System Engineering,
Kyungpook National University, Daegu,
Republic of Korea

Definition

Since the introduction of the first commercially available synthetic dye, *mauveine*, in 1856, traditional organic dyes have been used for the coloration of textiles by means of dyeing technology. The term “functional dye” was first coined by Japanese scholars in 1981 and stemmed from interest and activity in the field of dye chemistry interrelated with high-technology applications [1–6]. Nowadays, the term *functional dyes* refers to ultraviolet (UV) and particularly infrared (IR) active molecules which have been specifically designed for high-technology applications in fields such as electronic materials, electronic devices, reprographics, etc., as such “functional dyes” differ to traditional dyes.

Interactions of Functional Dyes

It is pertinent to examine the interactions of functional dyes with various agents before discussing the many and varied applications of functional dyes. Such interactions enable particular

phenomena to be associated with each application. Functional dyes are designed to interact with electromagnetic radiation, pH, electricity, heat, pressure, and even frictional forces.

Electromagnetic Radiation

Functional dyes interact with electromagnetic radiation in the near-UV (300–400 nm), visible (400–700 nm), and near-IR (700–1500 nm) regions to produce a variety of effects required for high-technology (hi-tech) applications. Selective absorption of nonvisible radiation, such as UV and particularly IR, is also important in applications such as optical data storage, computer to plate, security, and printing. A color change imparted by electromagnetic radiation is called *photochromism*. Photochromism is a chemical process in which a compound undergoes a reversible change between two states having separate absorption spectra, i.e., different colors. Although photochromism is undesirable in traditional textile dyes, it is widely used in, for example, spectacles and optical data storage. Indicator dyes are probably the most familiar example of a color-change effect.

Luminescence occurs when a molecule in an excited state, normally achieved by absorption of a photon, loses some or all of the excess energy as light rather than via the normal relaxation mode of heat loss. Luminescence from the first excited singlet state to the ground singlet state is known as *fluorescence*. Fluorescence is important in biological application, laser dyes, emissive displays, and in providing vivid, bright dyes, particularly for ink-jet printing.

Heat

Most molecules lose energy from the excited state as heat. The most efficient molecules for converting electromagnetic radiation into heat are those that absorb in the near-infrared region, namely, infrared absorbers (IRAs). IRAs have attracted much interest because of their use in laser thermal transfer, optical data storage, computer-to-plate printing, and as solar screens for car windscreens and windows.

Thermochromic dyes change color with temperature (heat). Thermochromic dyes find use in

Dye, Functional, Table 1 Application areas of functional dyes

Fields	Applications	
Data storage	Optical data storage	CD-R, DVD
	Printing	Ink for ink-jet printer, OPC, toner, thermal recording paper, D2T2
	Others	Pressure-sensitive paper, color shifting ink for counterfeit deterrence
Data display	Display	Color filter for LCD, organic EL, near-infrared absorption dyes for PDP
	Others	Digital papers, thermosensitive materials
Energy transfer	Solar cells	DSSC
	Others	Dye laser, nonlinear optical materials, organic semiconductor
Biomedical application		Diagnostic medical devices, laser dyes for photodynamic therapy

direct thermal printing and as temperature sensors, as well as in clothing and novelties.

Electricity

Functional dyes have been designed to interact with electricity to produce a color change or to fluoresce. *Electrochromic dyes* change their color, generally from colorless to colored, when an electrical voltage is applied. The main application areas for electrochromic systems are electrically switchable anti-dazzle rearview mirrors, glazing units for temperature and light control, as well as visual displays.

Frictional Force/Pressure

Piezochromic and *tribochromic functional dyes* have been designed to produce an electrostatic charge when mechanical means is applied, the former under applied frictional or grinding forces and the latter by pressure. Such compounds have not as yet enjoyed much commercial exploitation. Known as charge-control agents, *tribochromic* dyes are used in the toners for photocopiers and laser printers to both produce and regulate the triboelectric charge on the toner particles. Pressure-sensitive barochromic dyes find use in imaging (carbonless paper) and for testing pressure points in, for example, airplanes and cars.

Functional Dyes Classified by Application

As mentioned above, development of electronics industry and the affected industries is needed of

functional dyes, which are developed, discolored, or achromatized by very small energy such as light, heat, electricity, and so on. Functional dyes can be classified in several ways. This entry will describe specific examples of functional dyes by classified application such as data storage, data display, energy transfer, and so on. A list of fields and application of functional dyes are given in Table 1.

Data Storage

The data storage field can be divided into two main parts, namely, *optical data storage* and *printing*. The technologies related to printing seek to produce visible images on a hard copy, usually paper or special media; ink-jet, laser printing, photocopying, and thermal printing are examples of the printing techniques employed. However, there are other technologies which produce invisible images, usually for the entertainment and publishing fields, using functional dyes; *optical data storage* is such a technology which concerns invisible imaging.

In its broadest sense, optical data storage involves a medium which, by optical means, employs near-infrared radiation or visible light for the recording and retrieval of digital data. Various types of storage media have been developed, such as read-only compact discs (CD-ROM) to rewritable digital versatile discs (DVD-RW); optical data recording systems which use laser addressable organic dyes include WORM (write once read many), CD-R, and DVD-R.

The mechanism of optical recording using dyes involves converting excitation energy, which is generated by a focused laser beam, into thermal energy by non-radiative decay. In simple terms, a dye is coated onto a disc substrate to provide the recording layer. In writing mode, the laser is used near to full power and “blasts” holes in the very thin layer of dye at appropriate parts of the disc, as dictated by the digitally encoded information. When the information is retrieved, the power of the laser is turned down to avoid damaging the recording layer and is played over the surface of the disc to read the encoded signal with the aid of a detector. Since the mechanism of hole formation relies upon the dye absorbing laser radiation, heating up, and migrating from the hot region during the hole formation stage, the dye must absorb at the wavelength of the laser used. A number of technical criteria are demanded of suitable dyes, in addition to the general commercial requirements of low cost and high availability:

- Good solubility in the organic solvent used for spin coating the discs and the ability to form solid films.
- High stability for environmental and processing conditions, i.e., oxidation, hydrolysis, and light.
- High reflectivity of the solid film at the recording wavelength.
- Low thermal conductivity and hence high sensitivity for heat-mode recording.

The requirements for dyes in WORM, CD-R, and DVD-R media are very similar, but there are differences in their absorbance wavelength. Several types of infrared absorbers have received attention for use in optical data storage. Both phthalocyanine and cyanine colorants satisfy the commercially required technical criteria for optical data recording; phthalocyanines have many desirable attributes suitable for use in optical data storage.

Data Display

In recent times, there has been a significant development of display technologies which can

interface with an ever-increasing number of electronic devices. The role of the display is simply to render the output from the device in a form that is intelligible to the human reader; examples of devices include computers, calculators, watches, televisions, public signboards, etc.

Cathode Ray Tube

For many decades, the cathode ray tube (CRT) was the most common display used for televisions or visual display monitors. The operating principle in CRT involves an electron beam that is generated from a cathode located at one end of a glass vessel, which is held under vacuum. The electron beam is accelerated and is then moved horizontally or vertically using electrostatic or electromagnetic means. The inner, front surface of the glass vessel is coated with inorganic phosphors, and, when the beam of highly accelerated electrons hit this surface, the phosphors emit visible light. This process is known as *cathodoluminescence* which is an emissive technology and therefore produces bright images.

Liquid Crystal Display

Because of the large bulk and mass of CRTs, liquid crystal displays (LCDs) have been developed. LCDs employ specific organic molecules, such as *para*-cyanobiphenyl, which can be readily aligned in an electrical field. When the molecular axis of *para*-cyanobiphenyl is orthogonal to a beam of polarized light, the light is absorbed and a dark color results. When its molecular axis is parallel to the direction of the light beam, no absorption occurs and a pale-colored area results. This contrast allows monochrome imaging. Neither CRTs nor LCDs are the most suitable technology for very large area displays, where other display techniques such as plasma panels and electroluminescent devices offer advantage.

Organic Light-Emitting Devices

Organic light-emitting devices (OLEDs) are one of several technologies available for use as emissive flat panel displays. OLEDs possess the advantage that their molecular properties can be tailored, as well as inherent potential low cost, low

voltage operation, and relative simplicity of processing and fabrication into devices.

Two types of OLEDs are being developed: one based on multilayers of vapor-deposited, low molecular weight materials and other on polymeric materials, either by precursor polymerization or by casting from solution and subsequent processing. Most research has been done on low molecular weight material OLED.

A wide variety of low molecular mass materials have been designed for high-performance materials. The hole-transport layer (HTL) materials are normally substituted aromatic amines, which are known to have high hole mobility and which are applied to the surface of a glass substrate which has been coated with the conducting transparent material, indium tin oxide. The electron-transport layer (ETL) materials commonly used are emissive metal complexes, commonly aluminum, beryllium, or lanthanides. Some ETL materials are chosen because they are non-emissive so as to act as combined ET and hole-blocking layers.

Polymeric LED materials are a more recent development and have many attractive properties. One prime advantage is the ease with which it is possible to lay down thin films on large areas of a substrate either by spin coating or dip coating. Typical polymers are the green poly(*para*-phenylene vinylene) (PPV), the orange-red dialkoxy derivatives, and the blue polyfluorene.

Energy Transfer

Energy transfer via functional dyes involved in electronic materials has been discussed above. In this section, organic semiconductors and solar cells will be considered.

Electroluminescence (EL) involves the conversion of electrical energy into nonthermal light. To achieve this conversion, the materials can be either inorganic or organic. *Low-field* devices comprise conventional light-emitting diodes (LEDs), which are constructed from monocrystalline semiconducting materials.

Organic semiconductors are used in numerous devices, including field-effect transistors, light-emitting diodes, Schottky diodes, photovoltaics

(solar cells), light-emitting electroluminescent devices, etc. Organic semiconductor devices have been studied as unipolar devices based on hole-transport materials, because of the lack of suitable electron-transport materials. Conventional semiconductor LEDs are made by growing a region of *n*-type semiconductor material which has been doped with donors and *p*-type semiconductor materials doped with acceptors.

Organic Semiconductor

Organic semiconductors can be small molecules or polymers as employed in the case of electroluminescent molecules for emissive displays. Hole-transport materials include oligo- and polythiophenes, poly(thienylene vinylenes), and pentacenes. Electron-transport materials include copper phthalocyanine, hexadecafluoro copper phthalocyanine, naphthalenetetracarboxy dianhydride, and perylenetetracarboxy dianhydride.

Solar Cells

Solar cells have enjoyed much interest, especially in recent times. Attempts to mimic photosynthesis using chemicals closely related to chlorophyll such as phthalocyanine constitutes have received interest. However, the efficiency of most solar cells, including those based on silicon, is very low and requires improvement in order for solar cells to make a real impact.

Most recent research is focused on dye-sensitized solar cells (DSSCs) which provide both technically and economically viable alternatives to present-day photovoltaic devices. In DSSCs, the two processes of light absorption and charge carrier transport are separated, contrary to conventional systems. Light is absorbed by the *dye sensitizer* which is anchored to the surface of a wide-bandgap semiconductor. Charge separation takes place at the interface via photo-induced electron injection from the dye into the conduction band of the solid. Carriers are transported in the conduction band of the semiconductor to the *charge collector*. Relatively inexpensive semiconducting materials are used, such as titanium dioxide, zinc oxide, and tin oxide; various dyes are used.

DSSC technology has been used as an energy provider for devices using electron chromic cells in smart windows and display.

Nonlinear Optical Dyes

Nonlinear optics (NLO) is concerned with the interaction of electromagnetic radiation with various media to produce new radiation that is altered in phase, frequency, and amplitude from the incident radiation. Electrooptical activity, applying a field to change the refractive index of a material, means that the incoming light can be manipulated from on path to another.

While there are several types of nonlinear effects, the most important is *second-order frequency doubling*, whereby the incident radiation is converted to radiation of double the frequency; this has many applications in both telecommunications and optical data storage. For example, since in telecommunications the most efficient way to transmit data is via infrared laser radiation, consequently, an important use of nonlinear optical materials would be to convert the infrared radiation to visible radiation by frequency doubling, thus enabling easier detection of the signals. An example in optical data storage is that frequency doubling would enable more data to be stored per unit area because of the smaller spot size of the frequency-doubled radiation. Molecules capable of high polarizability give the best NLO effects; in this context, organic molecules having powerful donor and acceptor groups conjugated through a delocalized π -electron system give the largest second-order effects.

Laser Dyes

Although both solid-state and semiconducting lasers have made many inroads into applications that previously were the domain of dye lasers, nevertheless, dye lasers continue to offer several advantages:

- Tunability over a wide range of wavelengths from <400 to 1000 nm.
- Hyperfine tuning.
- High average power in both pulsed and continuous modes.
- High feed power.

- Ideal light source for the generation of short pulses.

Dyes have the important role in dye lasers of allowing a fixed wavelength of laser input to be tuned to a wide range from which a selection can be made appropriate for the desired end use. Modern laser dyes have changed little from those used a decade or so ago; coumarin and xanthene colorants are the two most important colorant types. Recently, dye lasers operating in the spectral region of 600–750 nm have been of particular interest as they have found applications in photodynamic therapy of cancer, medical diagnostics, surgery, dermatology, urology, fluorescent immunoassays, laser isotope separation, and environmental monitoring.

Biomedical Applications

Dyes have a long history of use in biomedical fields both for diagnostic and therapeutic purposes. Many dyes were used for staining bacteria and azo dyes can selectively detect metal cations. Recently, attention has focused on two interrelated aspects of *photomedicine*, namely, *phototherapy* and *photodiagnosis*. Phototherapy concerns the use of ultraviolet, visible white light, or near-infrared radiation to treat a variety of diseases, while photodiagnosis relates to the use of optical methods, based on fluorescence, for early diagnosis.

A vast range of organic chromophoric ring systems exhibit fluorescence. Fluorescence techniques are commonly used in biological, medical, and drug development. As such, fluorescent labels and probes have been employed in immunoassay, fluorescence microscopy, DNA sequencing, detection on arrays, flow cytometry, gel based quantification of nucleic acids and proteins, capillary electrophoresis, and single-molecule detection.

The *fluorophores* used in these analytical and biological outlets are based on cyanines, coumarins, naphthalimides, anthraquinones, acridines, etc.; however, such colorants have to be modified in order to be suitable for labeling purposes, DNA

fluorescence, etc. Cyanines, in particular, are the most popular and extensively researched colorant type for use in biological and medical imaging protocols owing to their excellent synthetic flexibility and fluorescence properties. Indeed, cyanines are the only synthetic, near-infrared fluorophores which are commercially available for use in this area.

Smart Textiles with Photochromic Dyes

A photochromic compound is characterized by its ability to undergo a reversible color change. Many photochromic organic compounds have been found and widely investigated in various fields. Spiropyrans and the closely related spirooxazines are important classes of photochromic dyes. While these two classes of compound are similar in many respects, the replacement of the benzopyran ring with a naphthoxazine ring results in spirooxazines having the advantage of greatly improved resistance to prolonged UV irradiation, which confers greater commercial importance to them. Applying photochromic dyes to a fabric can impart to the fabric “smart” fashion effects as well as enhanced UV protection. In this context, the application of spirooxazines to textiles has attracted research interest [7–11]. When UV light was applied to spirooxazine dyeings, photochromic reactions were mainly observed at 610 nm in which the color intensity of the dyeings increased with irradiation time. The color-produced photomerocyanine form of the colorant is usually unstable and returns to the original spiro form. The color strength (K/S) of the color-produced spirooxazine dyeings decreased with increasing storage.

Cross-References

- [Colorant, Nonlinear Optical](#)
- [Colorant, Photochromic](#)
- [Colorant, Thermochromic](#)
- [Dye: Photodynamic Therapy](#)

References

1. Hunger, K. (ed.): *Industrial Dyes: Chemistry, Properties. Applications.* Wiley-VCH, Weinheim (2004)
2. Bamfield, P. (ed.): *Chromic Phenomena: Technological Applications of Colour Chemistry.* The Royal Society of Chemistry, Cambridge (2001)
3. Waring, D.R., Hallas, G. (eds.): *The Chemistry and Application of Dyes.* Plenum, New York (1990)
4. Kim, S.H. (ed.): *Functional Dyes.* Elsevier, Amsterdam (2006)
5. Gregory, P.: *High-Technology Applications of Organic Colorants.* Plenum Press, New York (1991)
6. Yoshida, Z., Kitao, T.: *Chemistry of Functional Days.* Mita Press, Tokyo (1989)
7. El-Shishtawy, R.M., et al.: *Int J Photoenergy.* **21**, 434897 (2009)
8. Durr, H. (ed.): *Photochromism: Molecules and Systems.* Elsevier, Amsterdam (2003)
9. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds.* Plenum, New York (1999)
10. Lee, S.J., Son, Y.A., Suh, H.J., Lee, D.N., Kim, S.H.: *Dyes Pigments.* **69**, 18–21 (2006)
11. Son, Y.A., Park, Y.M., Park, S.Y., Shin, C.J.: *Dyes Pigments.* **73**, 76–80 (2007)

Dye, Ink-Jet

Algy Kazlauciusas

Department of Color Science, University of Leeds, Leeds, West Yorkshire, UK

Definition

An inkjet dye is a solvent-soluble colorant present in inkjet ink formulations to produce text and photo-realistic images for drop-on-demand piezo and thermal deskjet printers, signage for continuous industrial inkjet printers, and all manner of designs for piezo and thermal inkjet textile printers.

Principal Applications

Dyes used for inkjet printing can be placed into three categories:

- (a) Deskjet photo printing (A4 and A3 formats) of photo-realistic images, using drop-on-demand piezo and thermal inkjet systems

Paper substrate materials containing ink-specific ink-receivable layers are used, with both dye and pigment colorants being utilized.

Wide-format printing of images onto paper substrates (Giclée prints), using *Iris Graphics* printers, was the dominant method for producing large format photo-realistic fine art prints during the 1980s and early 1990s. These printers employed dye-based ink formulations to produce images. However, these printers have long been replaced by a range of wide-format printers capable of producing images on a range of substrate materials, including paper, poly(vinyl chloride), and poly(propylene). This new breed of printer type uses pigments as the colorant in all ink formulation types, i.e., water-based inks, solvent-based inks, and UV curing inks. The dyes which previously had been employed for large format images, especially for outdoor purposes, were replaced with pigments because of the inferior light fastness of the dye prints.

- (b) Industrial inkjet printing

This area is associated with continuous inkjet printing that is used for high-definition coding on a range of substrates, bar codes, micro printing, logos, special characters and graphics. For this type of inkjet printing, solvent dyes are heavily utilized. In addition, this category also encompasses piezo micro drop-on-demand and large character, valve jet technologies. For these latter types of printing, highly dispersed pigments are favored.

- (c) Printing of fabrics

Both dyes and pigments are used for the printing of textiles, with reactive dyes primarily used for cotton, viscose, and other cellulosic fabrics; acid dyes for silk, polyamide, and wool; disperse dyes for polyester; and pigments for a wide range of different fabrics.

Deskjet Photo Printing

In the early days of deskjet printing, prior to the advent of photo-realistic imaging, both thermal and piezo drop-on-demand printers used aqueous dye-based inks only. The first-generation dyes utilized were commercially available direct and acid dyes produced for the textile market as well as food dyes [1–5]. As prints produced using vibrant colors, as opposed to dull colors, are far more attractive to the average observer, it is no surprise that the early dyes used were chosen primarily for their vibrancy. Typical dyes chosen were C.I. Acid Yellow 23 (monoazo), commonly known as tartrazine, C.I. Acid Red 52 (xanthene), and C.I. Acid Blue 9 (triphenylmethane) with C.I. Food Black 2 (disazo) chosen as the black dye [6]. Unfortunately the vibrancy of these dyes was seriously offset by the poor light fastness of the ensuing prints. Subsequent dye introductions included azo H-acid dyes such as C.I. Reactive Red 180 for the magenta component, the phthalocyanine dye C.I. Direct Blue 199 for the cyan, C.I. Direct Yellow 132 as an alternative yellow, and metalized C.I. Reactive Black 31 for the black component. It should also be noted that for a number of deskjet ink sets, aimed at the photo-realistic imaging market, C.I. Pigment Black 7 (carbon black) was utilized instead of a black dye.

At the time, deficiencies in the inks including limited water solubility and insoluble impurities such as salts were addressed by replacing strongly acidic sulfonic acid groups with weakly acidic carboxylic acid groups so as to give high solubility in alkaline inks and by the use of dialysis/ultrafiltration purification techniques to remove insoluble impurities. At this point in time, the key properties required of inkjet dyes for photo-realistic imaging are brilliant/vibrant colors; high resolution; low color bleed (ideally none); high color strength; 5–20% solubility; very low salt content (chloride, sulfate, calcium, and iron all less than 50 ppm); good fastness to light, water, and smear; thermal and chemical stability; high surface tension; nontoxicity; and good storage stability at extremes of temperature.

Further key aspects in the drive to rival silver halide images were the introduction of light magenta and light cyan inks to improve skin tones and the gradual development over time of ink-substrate compatibility using ink-receivable layers, namely, swellable polymer (based on polyvinyl alcohol, gelatin, etc.) and pigmented microporous (silica/alumina), coated onto a conventional photographic resin-coated paper base [7–9].

A further development in the late 1990s was the introduction by Epson of the first pigment-based desktop inkjet printer. While this was a further attempt to improve the light fastness of photo-realistic images, other image attributes were sacrificed, namely, a reduced color gamut, the propensity to bronzing, and metamerism problems.

Currently deskjet, photo-realistic, inkjet printing principally uses anionic water-soluble dyes, incorporating sulfonic/carboxylic acid groups to ensure solvent solubility [10, 11]. With respect to Colour Index dye classification, these include acid dyes, direct dyes, and passivated reactive dyes [12]. The general approach, when used with the more commonly available microporous ink-receivable layer, is that the anionic dyes are attracted to the cationic polymers that modify the surface of the silica/alumina particles.

Table 1 highlights the more commonly used water-soluble dyes that are commercially available for photo-realistic deskjet printers [13, 14]. The table has been assembled with the aid of information gleaned from the Society of Dyers and Colourist's Colour Index and the web sites www.clariant.com, www.pigments.clariant.com, www.nip.clariant.com, and www.dyes.com.

Although pigment-ink photo-realistic printers now tend to dominate the A4/A3 market, due mostly to their light fastness, many photographers still prefer dye-ink prints rather than equivalent pigment-ink and conventional silver halide prints. When Clariant launched their Claria dye-based ink set in 2007, this revitalized the demand for dye-based photo-realistic images, the dyes [15–17] displaying light fastness ratings of at least 6 [18]. A comprehensive survey in 2008 (Zhang) revealed the preference of dye-based prints in the

key performance areas of color gamut, image resolution, and image contrast. Although dye-based inkjet prints were also perceived to be superior to pigment-based inkjet prints, regarding skin tone representation, they were perceived to be inferior to conventional silver halide prints [19].

A review of recent patent literature indicates ongoing research activity with respect to improved performance for dye colorants used for printed images, in the areas of improved image permanence on porous photo media [20, 21], or dye admixtures to extend chroma [22].

Industrial Inkjet Printing

Industrial inkjet printing is primarily utilized on production lines, or in manufacturing processes. Industrial inkjet printers can be broadly classified as continuous or drop on demand, although there are variants within each of these individual classifications.

Continuous printing technology [23] grew on the back of legislative requirements for product identification, i.e., product and data coding. Thus, this type of inkjet technology is commonly used to print the date, together with 1D and 2D bar codes, onto a range of different substrates, e.g., cartons, bottles, and cans. As this type of printer is used on moving production lines, the inks must be able to dry quickly; this is achieved using volatile solvent-based inks. Although the original inks employed were methyl ethyl ketone (MEK)-based, the drive for the minimization of volatile organic compounds (VOC's) has resulted in the introduction of solvent mixtures such as acetone in combination with lactate esters [24].

Although solvent dyes are predominantly used in solvent ink formulations, a number of black inks using carbon black pigment have been developed, specifically for bar code-OCR readability. A key advantage of using dyes for solvent-based inks is the fact that they are able to provide their own conductivity, provided either by a counter ion or via existing residual salts. In addition, dyes also have the advantage over pigments that they are present as a solution rather than as a dispersion. This means that the stability of the ink generally is

Dye, Ink-Jet, Table 1 Inkjet water-soluble dyes

Colour Index reference	Chemical class	Light fastness range	Additional information	C.I constitution number
C.I. Acid Yellow 3	Quinoline	3	Shading	47005
C.I. Acid Yellow 23	Monoazo	4–5	–	19140
C.I. Direct Yellow 86	Disazo	5	–	–
C.I. Direct Yellow 132	Disazo	4–5	High performance images	–
C.I. Direct Yellow 169	Disazo	4–5	–	–
C.I. Reactive Yellow 37	Monoazo	4	–	–
C. I. Reactive Red 23	Azo Cu complex	6	High performance images	16202
C. I. Reactive Red 24:1	Monoazo	3–4	–	–
C.I. Reactive Red 180	Monoazo	4	–	181055
C.I. Acid Red 52	Xanthene	2	Shading	45100
Undisclosed	Azo Cu complex			
Reactive magenta from BASF	6	–	–	
C.I. Acid Blue 9	Triarylamine	2–3	Shading	42090
C.I. Direct Blue 199	Phthalocyanine	7	High performance images	74190
C.I. Acid Black 94	Azo/1:2 Cr-complex	5–6	–	–
C.I. Direct Black 168	Trisazo	5	–	–
Undisclosed	Azo Cu-complex black from Clariant	5–6	High performance images	–
Undisclosed	Polyazo black from Clariant	5	–	–

n.b. Fastness properties according to DIN ISO 12040 using an aqueous-based ink-jet ink containing 4% black dye or 2.5% color dye onto plain paper

Clariant Brochure – products for nonimpact printing/colorants for water-based ink-jet inks

superior to an equivalent pigment-based ink, both in storage and when in use in the printhead.

Table 2 shows the most commonly used solvent dyes employed for industrial continuous inkjet printing on a range of different substrate materials. The table has been assembled with the help of information provided by the Society of Dyers and Colourist's Colour Index and a selection of web sites, i.e., www.nip.clariant.com, www.pigments.clariant.com, and www.dyes.com.

Inkjet Textile Printing

Whereas the 1990s were key years for the development of both deskjet and wide-format inkjet printing with respect to photo-quality images onto IRL-coated paper substrates [25], inkjet textile printing did not become a widespread

commercial proposition until 2003, when several inkjet industrial printers were launched at ITMA (International Textile Machinery Association), Birmingham, UK. One key aspect of these printers was the introduction of higher-resolution printheads.

While the market for digital textile printing continues to grow, the pace of growth is not dramatic, with digital printing, accounting for only 1% of the total output of textile printing technologies [26]. However, inkjet textile printing is suitable for short runs, small batches, and “just in time” production, making it ideal for upmarket textile items such as luxury home furnishings, fashion dress, and fashion accessories. As Italy is acknowledged as the major area for the production of high-end textile printing in Europe, a high percentage of commercial digital inkjet systems are located within this region. Although major

Dye, Ink-Jet, Table 2 Inkjet solvent dyes

Colour Index reference	Chemical class	Light fastness range	C.I. constitution number
C.I. Solvent Black 27	Azo Cr-complex	7	–
C.I. Solvent Black 28	Azo Cr-complex	7	–
C.I. Solvent Black 29	Azo Cr-complex	7	–
C.I. Solvent Black 45	Azo 1:2 Cr-complex	7	–
C.I. Solvent Blue 44	Phthalocyanine	6	–
C.I. Solvent Red 91	Azo 1:2 Co-complex	6	–
C.I. Solvent Red 127	Azo 1:2 Cr-complex	4	–
C.I. Solvent Yellow 83:1	Azo 1:2 Cr-complex	7–8	–

n.b. Fastness properties according to DIN ISO 12040 using a print containing 5 % dye in vinyl copolymer, applied at 2–3 g/m² on aluminum foil.

Clariant Brochure – products for nonimpact printing/colorants for solvent-based inkjet inks.

technological improvements have been claimed since ITMA 2007 in Munich, further improvements are required, namely, penetration of colorants into the textile substrate, a wider color gamut, the development of a universal ink for all textile substrates, specialty printing, and greater print production speeds [27].

The desired requirements for dye-based inks utilized in digital inkjet printing on textiles are not dissimilar to the requirements needed for inkjet photo-realistic imaging onto ink-receivable layer-coated papers. Key requirements are as follows:

- (a) Good optical density
- (b) Narrow absorption curve
- (c) Minimal color to color bleed
- (d) Good light fastness
- (e) Good water fastness
- (f) High solvent solubility
- (g) Minor levels of inorganic salts (less than 100 ppm chloride and sulfate anions)
- (h) Thermally and chemically stable

It is common practice that inkjet printing onto textile fabrics, using reactive and acid dyes, generally involves pretreatment and posttreatment to ensure dye fixation onto the fabric. There are a number of important reasons as to why the dyes are kept apart from the thickeners and other required chemicals, prior to fabric application:

- (a) Many of the chemicals required for the process could cause corrosion of the inkjet

printhead and/or instigate stability problems in the ink.

- (b) The rheological properties of the thickeners used are incompatible with the printheads
- (c) Reactive dyes may hydrolyze if alkali is present in the ink formulation.
- (d) Large levels of salt in aqueous inks would not only contribute to print-head corrosion but also reduce the level of dye solubility.

Water-based dye inks tend to be specifically used for piezo-electric and thermal printheads and usually contain a glycol solvent as a humectant. As previously mentioned, reactive dyes are used in ink formulations specifically prepared for the printing of cotton, viscose and other cellulosic fabrics. For wool, silk, and nylon fabrics, acid dyes tend to be the norm, whereas for polyester fabrics disperse dyes are used. Huntsman with their Novacron[®] reactive dyes, Lanoset[®] acid dyes, and Terasil[®] disperse dyes are a major supplier of digital inkjet dye-based inks. Other key companies involved in producing these inks are Dupont, with their Artistri range of dye-based inks, along with DyStar, Kiian, Sensient Technologies, and Kolorjet.

To ensure dye fixation, steaming under atmospheric pressure at just over 100 °C is required for both acid and reactive dyes. However, for the fixation of disperse dyes, a higher temperature of 170–180 °C is required, although this can be lowered to between 130 and 150 °C if pressure steaming is employed.

Other less popular methods of dye-based inkjet printing on textiles are still in use, namely, solvent-based, oil-based, and phase change. Hot-melt inks are involved in the phase change method, with the inks being solid at room temperature and fluid when heated to between 60 and 125 °C. Piezo printers tend to be used for this method of printing, and the ink vehicle is a mixture of fatty carboxylic acids and alcohol. Solvent dyes utilized are soluble in the vehicle and hence hydrophobic.

A wide range of research activity within the field of digital inkjet printing onto textile fabrics continues to be published, indicating a healthy future. The research areas involved include the development of both new and modified reactive dye structures for inkjet printing [28–30], fabric pretreatment [31, 32] and thickeners [33], development of inks using natural dyes [34], dye aggregation/interaction with water-soluble polymers [35], and new azo pyridone disperse dyes [36].

Cross-References

- [Colorant, Textile](#)
- [Coloration, Fastness](#)
- [Dye](#)
- [Dye, Metal Complex](#)

References

1. Kenyon, R.W.: Ink jet printing. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*, pp. 113–138. Blackie Academic and Professional, Glasgow (1996)
2. Wnek, W.J., Andreottola, M.A., Doll, P.F., Kelly, S.M.: Ink jet ink technology. In: Diamond, A.S., Weiss, D.S. (eds.) *Handbook of Imaging Materials*, 2nd edn, pp. 531–602. Marcel Dekker, New York (2002)
3. Bamfield, P., Hutchings, M.G.: Phenomena involving the absorption and reflection of light. In: *Chromic Phenomena, Technological Applications of Colour Chemistry*, 2nd edn, pp. 141–233. RSC Publishing, Cambridge (2010)
4. Bauer, W., Ritter, J.: Tailoring dyes for ink-jet applications. *Surf. Coat. Int.* **9**, 407–410 (1996)
5. Gregory, P.: Ink-jet printing. In: *High-Technology Applications of Organic Colorants*, pp. 175–205. Plenum Press, New York (1991)
6. Gregory, P.: Ink jet printing: joining the jet set. *Proc. RMS.* **36**, 232–238 (2001)
7. Kazlauciusas, A.: Longevity issues associated with photo-realistic ink jet images. Paper presented at the 3rd international conference on preservation and conservation issues related to digital printing and digital photography, Institute of Physics, London, 24–25 Apr 2006
8. Kazlauciusas, A.: Photorealistic ink-jet digital printing – factors influencing image quality, image stability and print durability. *Color. Technol.* **126**, 135–144 (2010)
9. Frenkel, M.: Tailoring substrates for inkjet printing. In: Magdassa, S. (ed.) *The Chemistry of Inkjet Inks*, pp. 73–122. World Scientific Publishing, Singapore (2010)
10. Gregory, P.: Colouring agents for non-impact printing – a survey. *Surf. Coat. Int. B Coat. Trans.* **85**, 9–17 (2002)
11. McFall, P.: Ink technologies for inkjet printing. *Surf. Coat. Int. A.* **10**, 420–426 (2004)
12. Soffel, J. L., Morris, P. C.: US6458195. Hewlett Packard Company (2002)
13. Carr, K.: Dyes for inkjet printing. In: Freeman, H.S., Peters, T. (eds.) *Colorants for on-Textile Applications*, pp. 1–34. Elsevier, Amsterdam (2000)
14. Fryberg, M.: Dyes for ink-jet printing. *Rev. Prog. Color.* **35**, 1–30 (2005)
15. Oki, Y., Kitamura, K.: EP 1700889. Seiko Epson Corporation. (2006)
16. Kitamura, K., Fukumota, H., Sao, A.: US20070263055. Seiko Epson Corporation. (2007)
17. Fukumota, H., Kitamura, K.: US20070263056. Seiko Epson Corporation. (2007)
18. Lei, Z.: Analysis and characterisation of inkjet dye-ink formulations, together with the longevity performance of various dye/ink substrate combinations. Leeds University, BSc Research Dissertation (2008)
19. Kazlauciusas, A., Zhang, M., Ball, J.: Investigation into digital print hard copy quality, longevity and durability. *J. Phys. Conf. Ser.* **231**, 012012 (2010)
20. Holloway, A.P., Sisk, M.W., Assignee, A.K.: US20110014375. Lexmark International. (2011)
21. Holloway, A.P., Sisk, M.W., Assignee, A.K.: US20110012968. Lexmark International. (2011)
22. Jackson, C., Kung, H.K., Valentine, J.E., Jackson, C., Kung, K.H.: WO 2009076457 A1, E.I. Du Pont de Nemours & Company (2010)
23. Kipphan, H.: *Handbook of Print Media*. Springer, Heidelberg (2001)
24. Samuel, J., Edwards, P.: Solvent-Based Inkjet Inks. In: Magdassi, S. (ed.) *The Chemistry of Inkjet Inks*, pp. 141–159. World Scientific, Singapore (2010)
25. Pond, S.: *Inkjet Technology*. Torrey Pines Research, California (2000)

26. Inkjet textile printing status report 2010. Paper presented at the Digital Inkjet Textile Seminar, Hangzhou (2010)
27. Ujji, H: State of art of inkjet textile printing – status report 2012. Paper presented at the 28th international conference on digital printing technologies, Quebec City, 9–13 Sept 2012
28. Clark, M., Yang, K., Lewis, D.M.: Modified 2,4-difluoro-5-chloro-pyrimidine dyes and their application in ink-jet printing on wool fabrics. *Color. Technol.* **125**, 184–190 (2009)
29. Maheshwari, A. C.: IN 2009MU00138 (2010)
30. Roentgen, G., Fekete, L., Nicollet, M.: WO 2012136428. Huntsman advanced materials. (2012)
31. Zhu, Q., Cao, J., Wei, W., Zhong, J., Yao, J., Ye, Y., Yang, X.: Effects of the cotton fabric pretreatment on application properties of digital inkjet printing with reactive dyes. *Adv. Mater. Res.* **331**, 398–401 (2011)
32. Liao, S.-K., Chen, H.-Y.: Ink-jet printing of nylon fabric using reactive dyestuff. *Color. Technol.* **127**, 390–395 (2011)
33. Yuen, C.W.M., Ku, S.K.A., Kan, C.W.: Use of a biomaterial as a thickener for textile ink-jet printing. *J. Appl. Polym. Sci.* **107**, 1057–1065 (2008)
34. Palacios-Guberti, S.B.: GB 2492641 (2013)
35. Park, J.-Y., Hirata, Y., Hamada, K.: Dye aggregation and interaction of dyes with a water-soluble poplymer in ink-jet ink for textiles. *Color. Technol.* **128**, 184–191 (2012)
36. Neubauer, S.: WO 2009118260 (2009)

Dye, Metal Complex

J. N. Chakraborty

Department of Textile Technology, National Institute of Technology, Jalandhar, Punjab, India

Synonyms

Premetallized acid dyes

Dye, Metal Complex,

Fig. 1 C.I. Reactive Violet 1

Definition

Metal-complex dyes and pigments are the coordination complexes of bi- or polyvalent transition metal ions with various dyes and have versatile application in various fields, which include the dyeing of nylon and protein fibers, paint, toners for photocopiers, laser and ink-jet printers, photoconductors for laser printers, nonlinear optics, singlet oxygen generators, dark oxidation catalysts, and high-density memory storage devices.

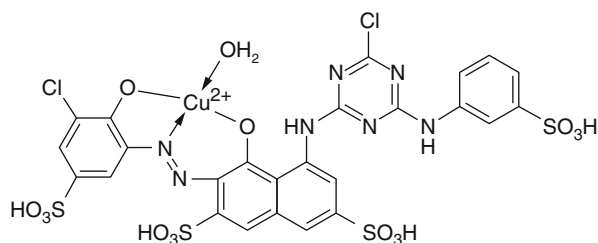
Chemistry and Application

Coloration of Textiles

The most prevalent examples are monoazo (C.I. Reactive Violet 1, Fig. 1) and disazo (C.I. Reactive Blue 82, Fig. 2) dyes. While formazan dyes are bi-, tri-, and tetradentate complexes, only tetradentate copper complex formazan dyes (e.g., C.I. Reactive Blue 160, Fig. 3) have found commercial use owing to their intense colors [1].

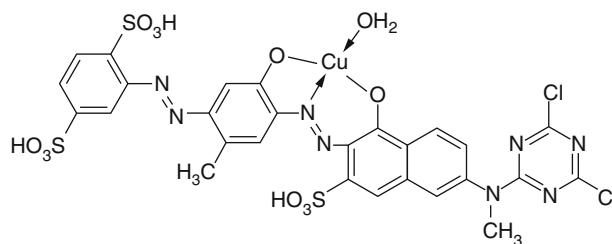
Metallization imparts a bathochromic shift producing deeper but duller shades. The spectra of 1:1 metal-complex azo dyes display narrower curves, while that of 1:2 complex dyes show a broader curve exhibiting blueness and dullness (Fig. 4).

1:1 metal-complex azo dyes are stable in strongly acidic pH and are preferably applied to wool, silk, and nylon fibers at a low pH (<4); they exhibit low to moderate coverage of fiber irregularities and poor compatibility in mixture. In contrast, 1:2 complexes are stable in weakly acidic pH and hence are applied from weakly acidic or neutral dyebaths (pH ~ 5–7); they display good buildup on wool, silk, and nylon fibers, exhibiting



Dye, Metal Complex,**Fig. 2** C.I. Reactive

Blue 82



C.I. Reactive Blue 71 (Fig. 12) are used in the coloration of paper and cotton, respectively.

Ink-Jet Printing

Ink-jet inks are water-soluble dyes in an aqueous vehicle. Metal-complex dyes and pigments possessing high absorbance, high chroma, and excellent lightfastness are employed. The copper complex dye C.I. Reactive Black 31 (Fig. 13) is very widely used as black in almost all ink-jet printers. For ink-jet printing of black on textiles, the 1:2 complex of a chelating azo dye with a mixture of metals “M” ($M = Cr^{3+}/Co^{3+}$, 70:30) is extensively used [8], as exemplified by C.I. Reactive Black 8 (Fig. 14).

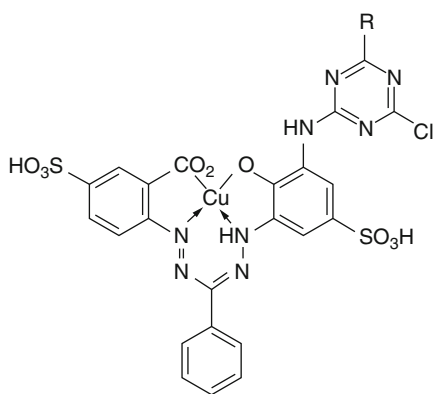
The black phthalocyanine dye (Fig. 15) was developed through reaction of amino phthalocyanine with acrylic or methacrylic acid, where “M” represents metal or hydrogen and “R” hydrogen or methyl. Solvent-soluble chromium-complexed azo black dyes, a mixture of 4- and 5-nitro, such as C.I. Solvent Black 35 (Fig. 16), is used in hot melt ink-jet printing [9].

Though yellow and magenta azo dyes are metal-free, the copper complex azo dye, C.I. Reactive Red 23 (Fig. 17), and the nickel complex PAQ dye (Fig. 18), where $R = H, Cl$, or $COOH$ and $R' = SO_2NHalkyl$, display dull and bright colors, respectively, each with similar high light fastness (Fig. 17) for ink-jet printers [10].

Copper phthalocyanine is extensively used as cyan for ink-jet printing; for textile ink-jet printing, metallized reactive dyes are used [8], such as C.I. Reactive Blue 71 (Fig. 12).

Photoconductors for Laser Printer

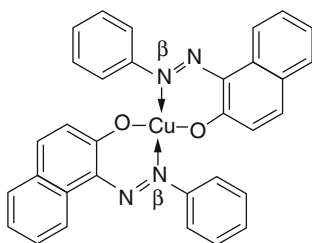
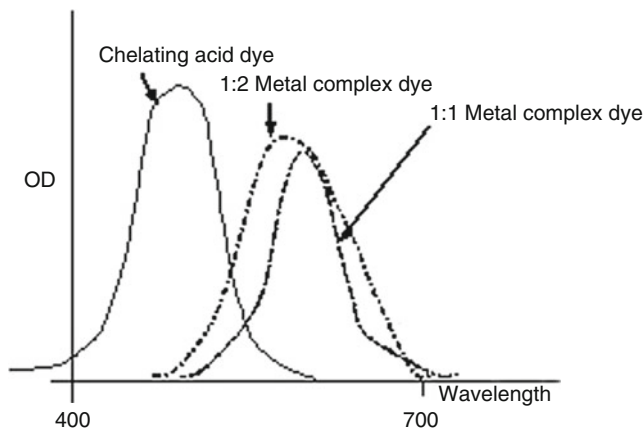
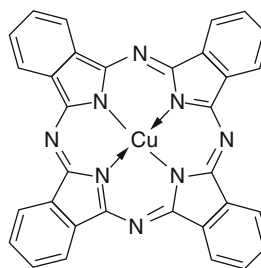
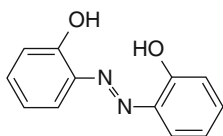
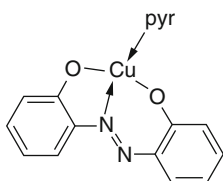
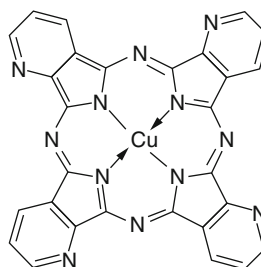
Photoconductors are dual-layer devices, and are exclusively used in laser printers. Metal-complex dyes, especially titanyloxy-phthalocyanine

**Dye, Metal Complex, Fig. 3** C.I. Reactive Blue 160

high compatibility in mixtures [2]. Dyeings of chromium complex dyes display high overall fastness compared to those obtained using copper, cobalt, and nickel complexes [3]. The metal remains bonded to the lone pair with one of the azo nitrogen atoms and not to the p-bond of the azo group in 1:2 copper complex dye (Fig. 5) [4]; however, in 1:1 copper complex dyes, the symmetrical dihydroxyazo ligand (Fig. 6) forms the 1:1 square planar complex (Fig. 7) [5].

Copper phthalocyanine (Fig. 8) is a bright blue pigment of λ_{max} 660nm characteristically displaying very high all-round fastness and it exists in several polymorphic forms [6, 7]. The redder, α -metastable form is preferred for use in paint, whereas the greener, stable β -polymorph is employed in printing inks. Complete chlorination of azo phthalocyanine blue pigment (Fig. 9) produces the high fastness green pigment, C.I. Pigment Green 7 (Fig. 10).

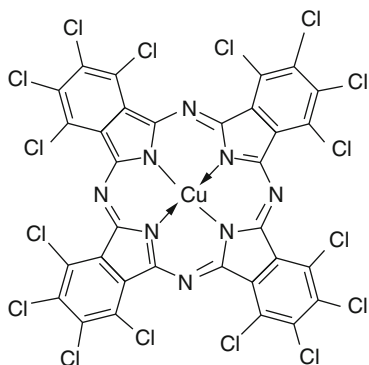
Water-soluble phthalocyanine dyes possess either sulfonic, sulfonamido, or both types of solubilizing group and are useful for the ink-jet printing of textiles. C.I. Direct Blue 199 (Fig. 11) and

Dye, Metal Complex,**Fig. 4** Spectral change in 1:1 and 1:2 complex dye with that of the parent dye**Dye, Metal Complex, Fig. 5** 1:2 Copper complex azo dye**Dye, Metal Complex, Fig. 8** Copper phthalocyanine (C.I. Pigment Blue 15:4)**Dye, Metal Complex, Fig. 6** 2,2'-Dihydroxyazobenzene (H₂DAB)**Dye, Metal Complex, Fig. 7** Copper (DAB). Pyr**Dye, Metal Complex, Fig. 9** Azo phthalocyanine blue pigment

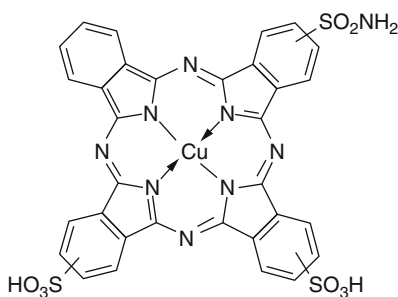
(Fig. 19), can act exclusively as charge generation materials (CGMs) in photoconductors [11]. All modern laser printers use titanyloxy-phthalocyanine, type IV polymorph, as the CGM possessing a “shallow shuttlecock” shape has a subtle effect on the packing of the molecules [12]. The pigment used as the photoconductor must be extremely pure and possess the correct morphology.

Toners for Photocopiers and Laser Printers

Laser printers, photocopiers, and especially ink-jet printers employ metal-complex dyes as toner. Laser printers and photocopiers use electricity and light to form an image with the help of a photoconductor; the latent electrostatic image on the photoconductor is thermally fixed using a toner [11, 12]. Metal-complex dyes and pigments produce the positive hole in the organic photoconductor and also perform key functions in

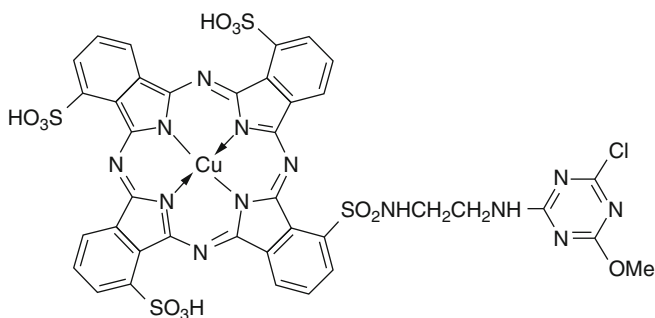


Dye, Metal Complex, Fig. 10 C.I. Pigment Green 7

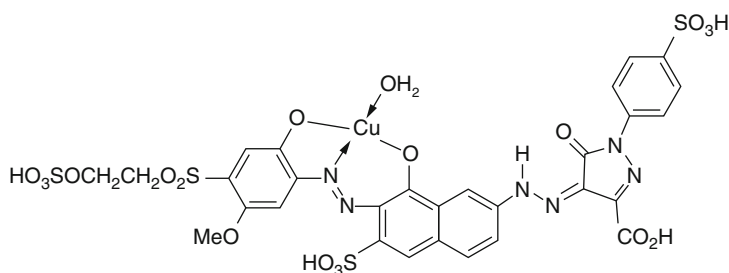


Dye, Metal Complex, Fig. 11 C.I. Direct Blue 199

Dye, Metal Complex, Fig. 12 C.I. Reactive Blue 71



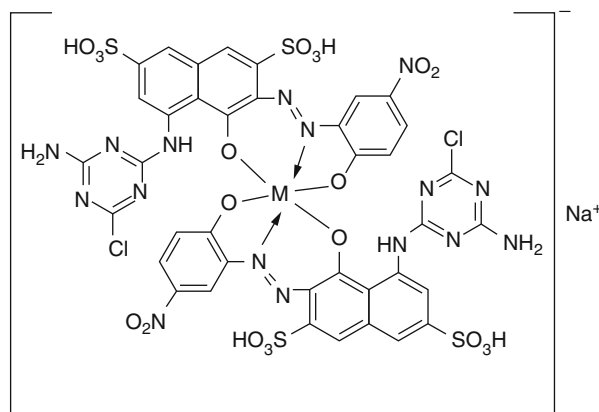
Dye, Metal Complex, Fig. 13 C.I. Reactive Black 31



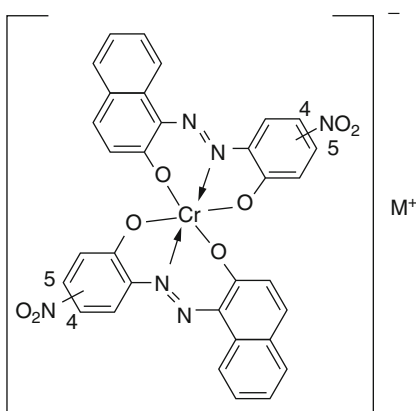
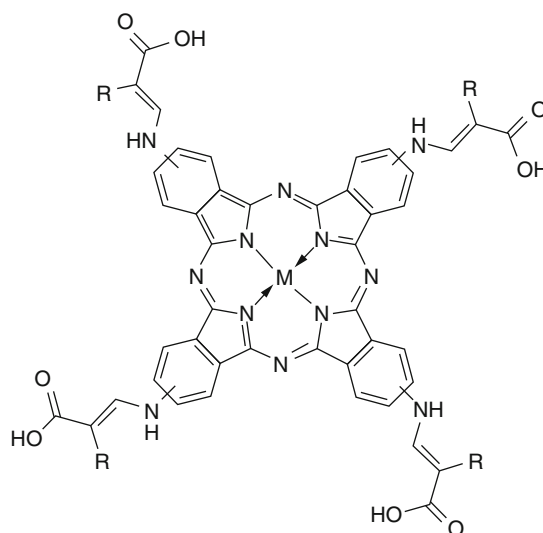
both the image generation stage and the toner development stage of the reproduction process.

Two main types of colored molecules are used in a toner: a colorant is employed to impart color to the toner, and the charge control agent (CCA) imparts color as well as control the triboelectric charge on the toner particles. Copper phthalocyanine, C.I. Pigment Blue 15:4 (Fig. 8), is the cyan colorant; however, pigments used for the yellow and magenta colorants are metal-free. CCAs can be either colored or colorless and are mainly metal-complex dyes. Colorless CCAs are useful for yellow, magenta, and cyan toners, while colored CCAs are more effective than colorless agents but are restricted to black toners. CCAs that impart a negative charge to the toner are invariably metal complexes, particularly 1:2 chromium-complexed azo dye with a mixture of 4- and 5-nitro (Fig. 16) and another one with a mixture of 4- and 5-chloro [12]. Efficient, colorless, negative-charging CCAs were developed by producing colorless analogues of 1:2 chromium complex of an aromatic *ortho*-hydroxy carboxylic acid, such as the complex of di-*tert*-butyl salicylic acid, e.g., *Bontron E-8136* (Fig. 20).

Another important class of supramolecular metal-complexed pigments is the so-called toner

Dye, Metal Complex,**Fig. 14** C.I. Reactive
Black 8

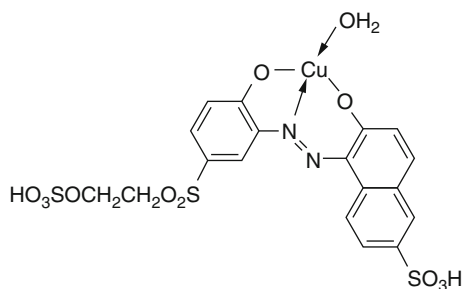
D

Dye, Metal Complex,**Fig. 15** Black
phthalocyanine dye**Dye, Metal Complex, Fig. 16** C.I. Solvent Black 35

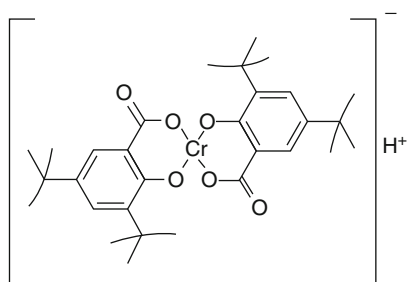
pigments which are water-soluble azo dyes containing sulfonic acid groups and are insolubilized by complexing with a divalent cation, such as calcium and barium. These compounds provide the standard magenta colorants for printing inks, such as C.I. Pigment Red 57:1, known as *Calcium 4B toner* (Fig. 21); chlorination of this product develops the red pigment which is suitable for characterization by synchrotron radiation.

Dark Oxidation Catalysts

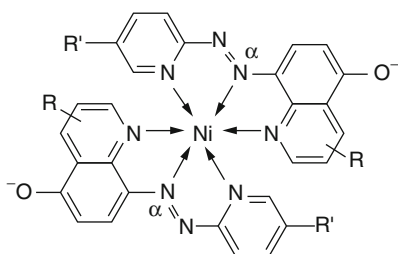
These are organometallic complexes based on transition metals possessing variable oxidation states and are used as **chemical transformers**



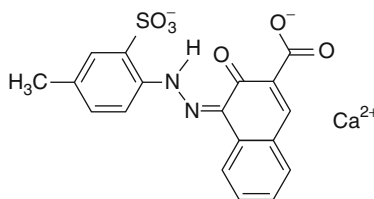
Dye, Metal Complex, Fig. 17 C.I. Reactive Red 23



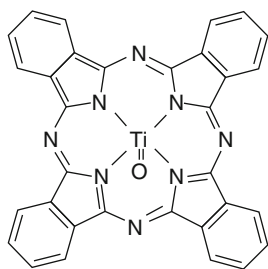
Dye, Metal Complex, Fig. 20 Bontron E-8136



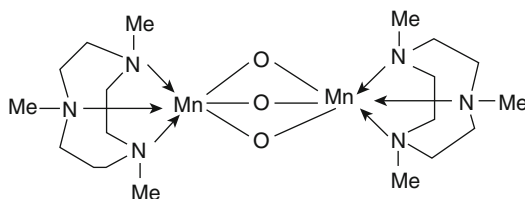
Dye, Metal Complex, Fig. 18 1:2 Nickel complex PAQ dye (PAQ: pyridyl azo Quinoline)



Dye, Metal Complex, Fig. 21 C.I. Pigment Red 57:1 (calcium 4B toner)



Dye, Metal Complex, Fig. 19 Titanyloxy phthalocyanine



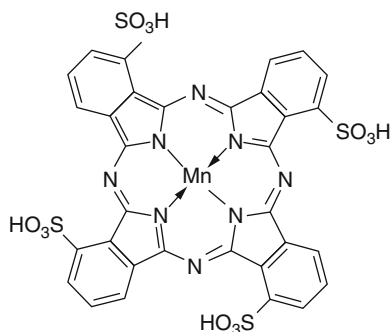
Dye, Metal Complex, Fig. 22 Organo-manganese complex (Unilever)

and in laundry powders [13]. Manganese complexes were developed and marketed, respectively, by Proctor and Gamble and Unilever, as washing powders [14–16]. Unilever's manganese complex (Fig. 22) was effective in removing dirt and stains from soiled fabrics, but this was accompanied by serious damage of the fabric. Proctor and Gamble developed a similar dark oxidation catalyst, manganese phthalocyanine tetra-sulfonic acid (Fig. 23), but could not achieve commercial success because of the “manganese factor” which

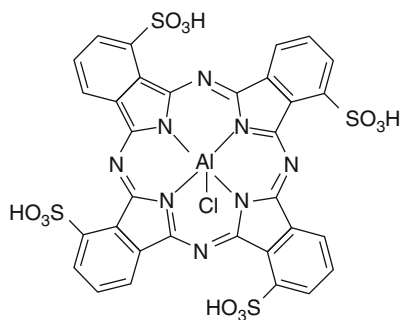
was probably at the root of disintegration of the fabric.

Singlet Oxygen Generators

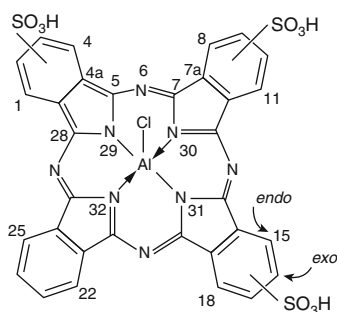
These have multipurpose applications, such as in skin, breast, and oropharyngeal cancer treatment through photodynamic therapy (PDT) using a photosensitizer [17, 18]. In photodynamic therapy, electromagnetic radiation in the visible and adjacent ultraviolet and infrared regions is used for photochemical treatment of (i) jaundice (neonatal hyperbilirubinemia) in the newborn, and especially in the prematurely born, (ii) psoriasis using light in the UV-A range and an administered photosensitizer, (iii) wet form of



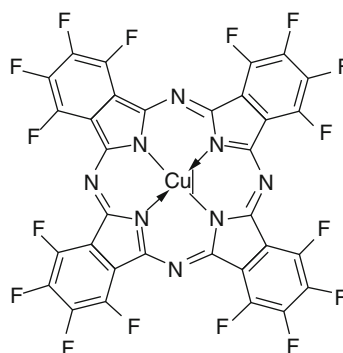
Dye, Metal Complex, Fig. 23 Manganese phthalocyanine tetra-sulfonic acid (Proctor and Gamble)



Dye, Metal Complex, Fig. 25 Tinolux BBS



Dye, Metal Complex, Fig. 24 Sulfonated of chloroaluminum (III) phthalocyanine (Photosens)



Dye, Metal Complex, Fig. 26 Hexa-deca-fluoro copper phthalocyanine

age-related macular degeneration with a photosensitizer as well as a laser light source, and (iv) certain cancer, such as skin, breast, and oropharyngeal cancer with a photosensitizer and red light [19–24].

The phthalocyanine-based commercial product *Photosens* (Fig. 24) is a widely used photosensitizer prepared by the direct sulfonation of chloroaluminum (III) phthalocyanine. Singlet oxygen generators are also used as hard-surface disinfectants, soaps, washing powders [25], and insecticides [26]. These eco-friendly photo-bleaches convert triplet oxygen ($^3\text{O}_2$) into highly reactive singlet oxygen ($^1\text{O}_2$) in the presence of light, thus acting swiftly to bleach stains. The blue aluminum phthalocyanine dye, *Tinolux BBS* (Fig. 25), is a commercial singlet oxygen generator [27]; zinc phthalocyanines, which absorb in the near-infrared region at ~ 725 nm, are also very good singlet oxygen generators [28].

Organic Semiconductors

Organic semiconductors are used to fabricate electronic devices. Metal-complex pigments, such as hexa-deca-fluoro copper phthalocyanine (Fig. 26), are used for electron transport [28].

Infrared Absorbers

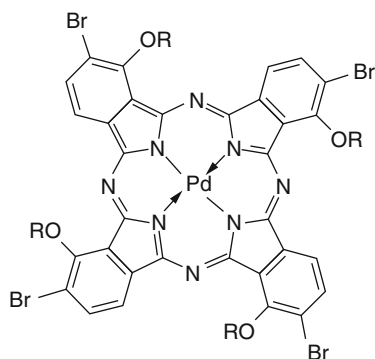
Upon complete chlorination, blue copper phthalocyanine is converted into its green-colored, fast pigment which also is an efficient infrared absorber, possessing high solubility in organic solvents and displaying narrow absorption curves at around 780 nm. Phthalocyanines containing amino donors, in combination with either thiols [29] or halogens [30], provide broader curves absorbing at around 900 nm.

Copper phthalocyanine loses its planar structure upon conversion to infrared absorber, with eight of the sterically demanding thiophenol groups lying above the plane of the

phthalocyanine ring and eight lying below. Taking advantage of this feature, non-planar copper phthalocyanines are used as high-level security components in currency notes [31, 32], in computer-to-plate technology to convert near-infrared radiation into heat [33, 34] and as solar screens on car windscreens and windows of buildings which allow daylight through but screen out infrared radiation component which generates heat [35]. However, as the durability of non-planar copper phthalocyanine is not high, palladium phthalocyanine (Fig. 27) is the leading phthalocyanine infrared absorber for CD-R applications [35, 36].

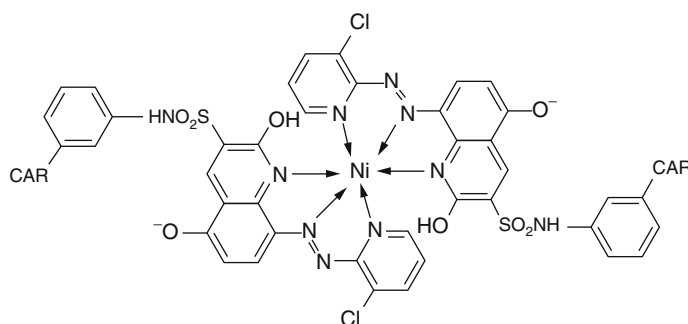
Silver Halide Photography

Metal-complex azo dyes are used in color diffusion transfer photography, employing non-diffusible, magenta dye-releasing colorants which, upon development of the silver halide layer, release a diffusible magenta dye; the nickel



Dye, Metal Complex, Fig. 27 Brominated tetra (α-2,4-dimethyl-3-pentoxy) palladium phthalocyanine

Dye, Metal Complex, Fig. 28 Nickel complex PAQ dye retaining CAR (Ballasted carrier moiety)



complex PAQ dye (Fig. 28), where “CAR” denotes Ballast-Carrier-Link, is a typical example [37]. Other metal-complex azo dyes include pyridylazonaphthols [38] and pyridylazoaminophenols [39].

Nonlinear Optics

Metal complexes, mostly of δ -transition metals, can exhibit nonlinear optical (NLO) behavior which can be used to manipulate the fundamental properties of laser light beams, thereby facilitating optical data processing and storage [40–42], and offering possibilities for the combination of quadratic (second-order) and cubic (third-order) NLO properties along with other molecular electronic properties. Examples include bimetallic complexes containing pyridyl carbonyl centers with ruthenium σ -acetylide; ferrocenyl (Fc) electron donor groups can also exhibit NLO properties [43–45].

In metalloporphyrin derivatives, such as metallophthalocyanines and metallo-2,3-naphthalocyanines, metallation greatly enhances the cubic responses of the compound; such materials often exhibit strongly absorbing, long-lived triplet excited states. Asymmetrically substituted metalloporphyrin derivatives can exhibit extremely large quadratic NLO responses; the β -response of the Cu(II) complex shows a marked wavelength dependence, while its Zn(II) analogue displays essentially identical β 830 and β 1064 values [46].

Some porphyrin-like metal complexes, mostly of Cd(II), in which one of the pyrrole rings is replaced with an aryl or vinyl group, exhibit very large cubic NLO responses [47, 48].

High-Density Memory Storage

The electrochromism exhibited by metal-complex dyes in the near-infrared region (NIR) of the spectrum (ca. 800–2,000 nm), obtained by the generation of different electronic absorption bands, enjoys applications such as optical data storage (ODS) on various disk media, i.e., read-only, write-once, and rewriteable media [49, 50]. In such applications, a layer of a metal-complex dye based on highly conjugated ligands displays a significant absorption maximum which is either at or very close to the wavelength of the laser light and also has relatively low optical absorbance and high reflectance at the recording wavelength and is interposed between the substrate and the reflecting layer of a CD. With appropriate optimization of the optical properties, this composition will record and reproduce [35, 51, 52].

Though polymethine, porphyrin, triphenylmethane, and azoic metal-complex dyes possess good photostability and absorption in the range 750–850 nm exhibiting electrochromism, phthalocyanine (Fig. 29), naphthalocyanines, and cyanine dyes based on indole and benzindole are the most widely used for the mass production of on-disk-structure (ODS) media because of the simplicity of modification, excellent photostability, and cost-effectiveness. However, selective metal-complex dyes should meet certain requirements; partial bromination allows fine-tuning of the film absorbance and reflectivity [35, 49, 53], commercially known as *Supergreen*. Phthalocyanine dyes are not suitable for DVD-R media, since the main chromophore cannot readily be modified to produce a sufficiently large hypsochromic shift; dyes potentially suitable for DVD-R include metal azo complexes, quinophthalones, and

diphenylmethanes. The cyanine dyes are very useful for all current optical disk recording applications [35].

A novel rewriteable optical disk comprising a Si-naphthalocyanine dye and the conductive polymer polythiophene as the recording layer has been reported which could be played back on conventional CD players, even after ten cycles of rewriting [54]. The phthalocyanine derivative has recently been reported to be capable of recording high-quality signals at high speeds [55].

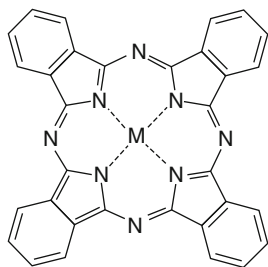
A series of metallized thiazolyl-azo dyes with a hydroxy substituent ortho to the azo linkage shows apparently promising performance [56, 57]. Zn(II) complex dyes possess solubility, thermal stability, and suitable absorption at 635 nm and are used as DVD-R dyes [58].

Cross-References

- [Coloration, Textile](#)
- [Colorant, Nonlinear Optical](#)
- [Coloration, Fastness](#)
- [Dye: Photodynamic Therapy](#)
- [Dye, Ink-Jet](#)

References

1. Gregory, P.: Chapter 9.12. In: McCleverty, J.A., Meyer, T.J. (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, vol. 9, 2nd edn. Elsevier, Oxford (2005)
2. Sekar, N.: *Colourage* **47**(4), 27–28 (2000). 35
3. Hannemann, K.: *Int. Dyer.* **186**(3), 16–19 (2001)
4. Price, R.: Dyes and pigments. In: Wilkinson, G. (ed.) *Comprehensive Coordination Chemistry*, pp. 35–94. Pergamon, Oxford (1987)
5. Adams, H., Bucknall, R.M., Fenton, D.E., Garcia, M., Oaks, J.: *Polyhedron*. **17**(23–24), 4169–4177 (1998)
6. Lenoir, J.: Organic pigments. In: Venkataraman, K. (ed.) *The Chemistry of Synthetic Dyes*, vol. V, pp. 313–474. Academic, New York (1971)
7. Gordon, P.F., Gregory, P.: *Organic Chemistry in Color*, pp. 219–226. Springer, Berlin (1983) Chapter 5
8. Provost, J., Gregory, P. (Avecia Ltd.): Multicolor Ink Jet Printing. US Patent 6,336,721, 2002
9. Kayaku, N.: World Patent Application, 200064901A, 1999
10. Evans, S., Weber, H.: Metal Complex for Ink Jet. US Patent. **6**, 001,161 (1999)



Dye, Metal Complex, Fig. 29 Metal Phthalocyanine

11. Gairns, R.S.: Electrophotography. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*, pp. 76–112. Blackie, London (1996)
12. Gregory, P.: *High-Technology Applications of Organic Colorants*, pp. 59–122x. Plenum, New York (1991) Chapter 7
13. Moser, F.H., Thomas, A.L.: *Phthalocyanine Compounds*. Rheinhold, New York (1963). pp 62, 109, 133, 134
14. Hage, R., Iburg, J.E., Kerschener, J., Koek, J.H., Lempers, E.L.M., Martens, R.J., Racherla, U.S., Russell, S.W., Swarthoff, T., Van Vliet, M.R.: *Nature*. **369**(6482), 637–639 (1994)
15. Martens, R.J., Swarthoff, T.: (Unilever): Detergent compositions containing dinuclear manganese complex as bleach. Canadian Patent. **2,083,661** (1993)
16. Gregory, P., Reynolds, S., White, R.L. (Zeneca): Sulfonated manganese phthalocyanines and compositions thereof. US Patent 5484,915, 1996
17. Dougherty, T.J., Gomer, C.J., Henderson, B.W., Jori, G., Kessel, D., Korbelik, M., Moan, J., Peng, Q.J.: *Natl. Cancer Inst.* **90**, 889–905 (1998)
18. Stranadko, E.F., Skobelkin, O.K., Vorozhtsov, G.N., Mironov, A.F., Beshleul, S.E., Markitchev, N.A., Riabov, M.V.: *Proc. SPIE*. **3191**, 253–262 (1997)
19. Bonnett, R.: Chapter 9.22. Metal complexes for photodynamic therapy. In: McCleverty, J.A., Meyer, T.J. (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, 2nd edn., vol. 9. Ward, M.D., (ed.) *Applications of Coordination Chemistry*. Elsevier, Oxford (2005)
20. Finsen, N.R.: Phototherapy; Sequeira, J.H., Transl. Arnold, London (1901)
21. Regan, J.D., Parrish, J.A. (eds.): *The Science of Photomedicine*. Plenum, New York (1982)
22. Lim, H.W., Soter, N.A. (eds.): *Clinical Photomedicine*. Marcel Dekker, New York (1993)
23. Bonnett, R.: *Rev. Contemp. Pharmacother.* **10**, 1–17 (1999)
24. Boehncke, W.-H.: *Curr. Opin. Anti-inflamm. Immunomodulatory Invest. Drugs*. **1**, 338–352 (1999)
25. Gregory, P., Thetford, D. (Zeneca): Poly-substituted phthalocyanines. US Patent 5486,274, 1996
26. Service, R.F.: *Science*. **268**, 806 (1995)
27. Tinolux BBS Trade Literature, Edition; Clba-Geigy, April, 1986. 578 Metal Complexes as Speciality Dyes and Pigments (1986)
28. Bao, Z.N., Rogers, J.A., Katz, H.E.J.: *Mater. Chem.* **9**, 1895–1904 (1999)
29. Stark, W.M.(ICI): Substituted Phthalocyanine. US Patent 4,824,947, 1989
30. Kaeida, O., et al. (Nippon Shokubai): Novel phthalocyanine compounds: production method thereof: near infrared ray absorption materials containing the same. European Patent 523,959, July 14, 1992
31. Gregory, P.J.: Porphyrins Phthalocyanines. **3**, 468–476 (1999)
32. Gregory, P.: *High-Technology Applications of Organic Colorants*, pp. 215–254. Plenum, New York (1991) Chapter 11
33. Haley, N.F., Corbiere, S.L.: Radiation-sensitive composition containing a resole resin and a Novolac resin and use thereof in lithographic printing plates. US Patent. **5**, 372,907 (1994)
34. Haley, N.F., Corbiere, S.L.: Method of making a lithographic printing plate containing a resole resin and a Novolac resin in the radiation sensitive layer. US Patent. **5**, 372,915 (1994)
35. Hurditch, R.: *Dyes for optical storage disc media*. Chemichromics '99. New Orleans. (1999)
36. Wolleb, H. (C.I.ba-Geigy): Process for the preparation of brominated, alkoxy-substituted metal phthalocyanines. US Patent 5,594,128, 1997
37. Evans, S., Elwood, J.K. (Eastman Kodak Co.): Photographic products and processes employing novel non-diffusible magenta dye-releasing compounds and precursors thereof. US Patent 4,420,550, 1983
38. Chapman, D., E-Ming, W. (Eastman Kodak Co.): Photographic products and processes employing novel nondiffusible heterocyclazonaphthol dye-releasing compounds. US Patent 4,207,104, 1980
39. Anderson, R.B., et al. (Eastman Kodak Co.): U.S. Patent Application Serial No. 282,613, 1981
40. Harper, P., Wherrett, B. (eds.): *Nonlinear Optics*. Academic, New York (1977)
41. Shen, Y.R.: *The Principles of Nonlinear Optics*. Wiley, New York (1984)
42. Saleh, B.E.A., Teich, M.C.: *Fundamentals of Photonics*. Wiley, New York (1991)
43. Calabrese, J.C., Tam, W.: *Chem. Phys. Lett.* **133**, 244–245 (1987)
44. Mata, J., Uriel, S., Peris, E., Llusar, R., Houbrechts, S., Persoons, A.J.: *Organomet. Chem.* **562**, 197–202 (1998)
45. Senechal, K., Maury, O., Le Bozec, H., Ledoux, I., Zyss, J.J.: *Am. Chem. Soc.* **124**, 4560–4561 (2002)
46. LeCours, S.M., Guan, H.W., DiMaggio, S.G., Wang, C.H., Therien, M.J.J.: *Am. Chem. Soc.* **118**, 1497–1503 (1996)
47. Gong, Q.H., Wang, Y., Yang, S.C., Xia, Z.-J., Zou, Y.H., Sun, W.F., Dong, S.M., Wang, D.Y.: *J. v Phys. D: Appl. Phys.* **27**, 911–913 (1994)
48. Sun, W.F., Byeon, C.C., Lawson, C.M., Gray, G.M., Wang, D.-Y.: *Appl. Phys. Lett.* **77**, 1759–1761 (2000)
49. Hurditch, R.: *Adv.: Color SC.I. Technol.* **4**, 33–40 (2001)
50. Hunt, P.A.: Optical data-storage systems. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*. Blackie, London (1996)
51. Hamada, E., Shin, Y., Ishiguro, T.: *Proc. SPIE*. **1078**, 80 (1989)
52. Mortimer, R.J., Rowley, N.M.: Chapter 9.13. In: McCleverty, J.A., Meyer, T.J. (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, 2nd edn., vol. 9. Ward M.D. (ed.)

- Applications of Coordination Chemistry. Elsevier, Oxford (2005)
53. Clba, Supergreen Product Dossier, Issue 2.0 (1999)
 54. Tomiyama, T., Watanabe, I., Kuwano, A., Habiro, K., Takane, N., Yamada, M.: *Appl. Opt.* **34**, 8201–8208 (1995)
 55. Verhoeven, J.A.Th., Mischke, W.S.: Developments in CD-R. In: Gan, F., Hou, L. (eds.) *Proc SPIE 2001. International Society for optics and photonix (SPIE). Fifth International Symposium on Optical Storage (ISOS 2000)*, vol. 4085, Bellingham, US, pp. 298–305
 56. Wang, S., Shen, S., Xu, H., Gu, D., Yin, J., Tang, X.: *Dyes Pigments*. **42**, 173–177 (1999)
 57. Wang, S., Shen, S., Xu, H., Gu, D., Yin, J., Dong, X.: *Mater. Sci. Eng.* **B79**, 45–48 (2001)
 58. Park, H.Y., Lee, N.H., Je, J.T., Min, K.S., Young, J.H., Kim, E.R., Lee, H.: *Mol. Cryst. Liq. Cryst.* **371**, 305–308 (2001)

Dye: Photodynamic Therapy

Mark Wainwright

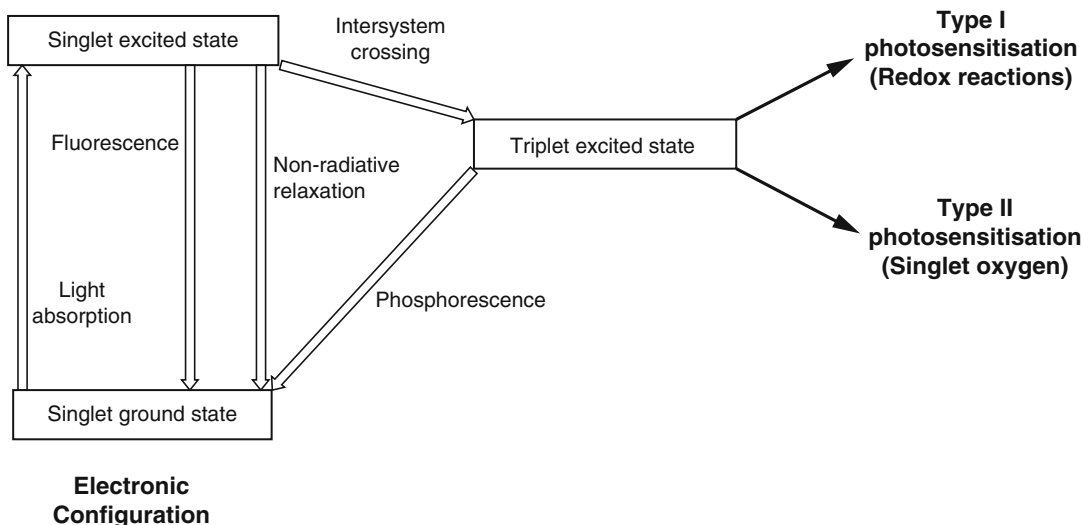
Pharmacy and Biomolecular Sciences, Liverpool
John Moores University, Liverpool, UK

Definition

Photodynamic therapy may be thought of as a directed and controlled application of the cellular

damage possible from light energy. There are two main areas of clinical interest, these being the *photodynamic therapy* of cancer (shortened to PDT) and the parallel application to microbial disease and infection control, the latter having several names, including *photodynamic antimicrobial chemotherapy* (PACT, used in this review), *antimicrobial photodynamic therapy* (APDT), and *photodynamic disinfection* (PDD). For both areas of clinical interest, in order to target and mediate cellular damage, a light-absorbing chemical, called a *photosensitizer*, is employed.

In both the anticancer and antimicrobial guises, the underlying science is the same, namely, (i) uptake by or association of the target cell with a photosensitizer, (ii) illumination of the target cell and the concomitant electronic excitation of the photosensitizer, and (iii) chemical reaction of the photosensitizer with its biomolecular environment or energy transfer from the photosensitizer to a local oxygen molecule. In both of these final scenarios, reactive oxygen species result which are highly toxic in the cellular environment (Fig. 1). The combination of locally administered photosensitizer and targeted delivery of illumination results in the production of a rapid, destructive



Dye: Photodynamic Therapy, Fig. 1 Photosensitization pathways

local effect by means of which local therapeutic control can be maintained that is absent of side effects in parts of the body remote from the treatment site.

Introduction

The interaction of light and matter is an ordinary, everyday occurrence, so much so that it is ignored by the majority. However, without the effects caused by solar radiation on plant species, there would be no life on planet Earth. Despite the perceived simplicity of plant life, photosynthesis, the process by which plants absorb light and store the energy in the form of carbohydrates, is a complex sequence of chemical interaction and energy transfer. This process is, of course, a positive use of light energy. The converse may be seen in the increasing incidence of skin cancers, usually in those individuals who, unwittingly or not, allow parts of their skin structure to be damaged by the higher energy portion of sunlight or by the radiation provided by sunbeds.

Photosensitizers and Photodynamic Therapy (PDT)

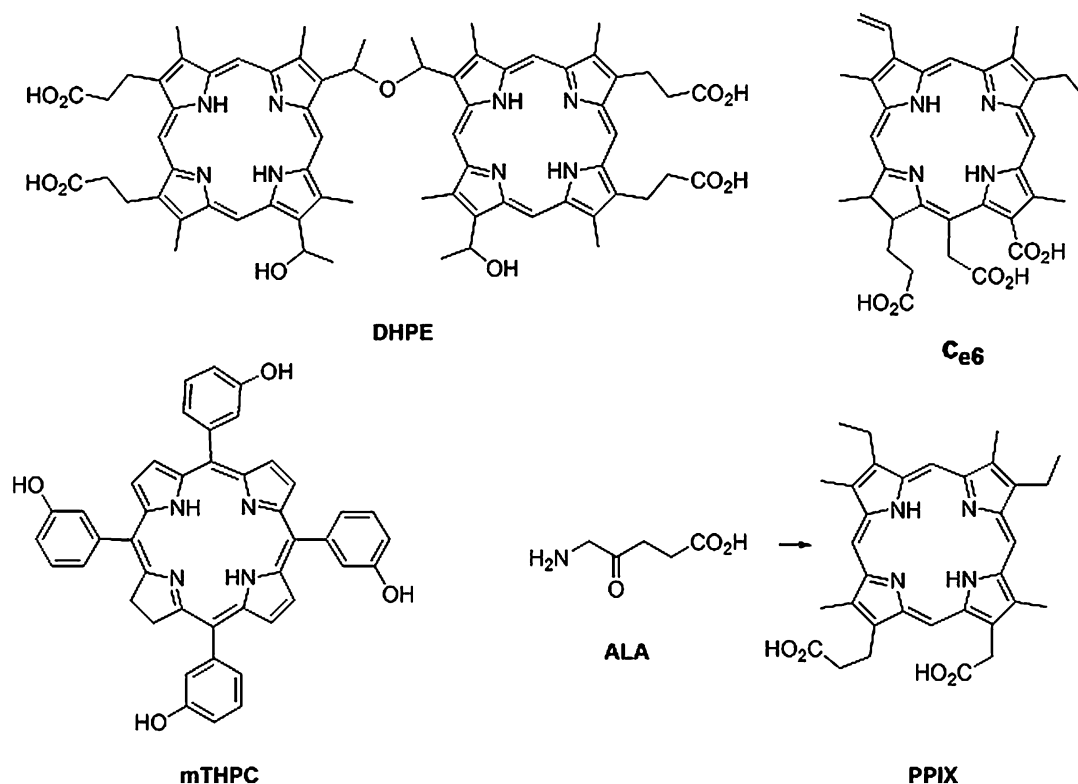
Ideally, PDT involves the use of a nontoxic photosensitizer which produces a damaging effect in the target cells upon illumination. In practice, this has proved elusive, because of either inherent or “dark” toxicity or lack of selectivity for target cells. However, in comparison to conventional chemotherapy or radiotherapy, PDT offers considerable advantages, mainly due to any side effects experienced being localized to the area of application. Compared to surgical excision, the demonstrable advantage of the PDT approach lies in much improved cosmesis, i.e., a lack of tissue damage, disfigurement, and scarring. This can be particularly important, for example, in head and neck cancers, from an aesthetic point of view, and also in external gynecological cancers due to loss of functionality and impaired sex drive.

Photosensitizers Used in PDT

Much of the original work which suggested photodynamic therapy as an approach to malignant disease was due to clinical findings, for example, Policard’s observation of the fluorescence of porphyrins in the blood supply of tumors under fluorescent light during surgery [1]. The porphyrin species involved were naturally occurring species, normally present in the blood. Later work, by Lipson and others, utilized porphyrin compounds derived from bovine blood and thus known as hematoporphyrins or hematoporphyrin derivatives (HpD; see Fig. 2). These were mixtures of short chains (oligomers) of hematoporphyrins (Fig. 2), and their constituency was variable by means of chemical derivation. However, despite the lack of purity entailed, HpD was shown to be reasonably effective in “shutting down” tumors via oxygen starvation, and purer forms, such as “porfimer sodium,” have been licensed clinically for a range of presentations. The main side effect associated with such photosensitizers derives from a lack of selectivity when administered systemically, which has often led to generalized patient photosensitization – i.e., a strong reaction to normal daylight, posttreatment.

So-called “second-generation” photosensitizers related to porphyrins for use in PDT include the chlorins mTHPC and chlorin e_6 (Fig. 2). Both of these agents offer the advantage of longer-wavelength light absorption than their predecessors, due to the marginally altered ring system (one carbon-carbon double bond is missing). However, nonoptimal tumor selectivity was again identified as a problem (with consequent generalized photosensitization as recounted above), and the search for improved photosensitizers has continued, producing new compounds related to the original porphyrins, but with variations on the size and constitution of the ring system [2].

Understanding of the heme biosynthetic pathway has led to the use of a natural porphyrin precursor, 5-aminolaevulinic acid (ALA, Fig. 2), and its derivatives in PDT. ALA is normally utilized in mammalian cell heme (protoporphyrin IX, PPIX, Fig. 2) synthesis, but the increased



Dye: Photodynamic Therapy, Fig. 2 Porphyrin photosensitizers used in PDT. *DHPE* dihematoporphyrin ether, a constituent of HpD and porfimer sodium; *ce6* chlorin e₆;

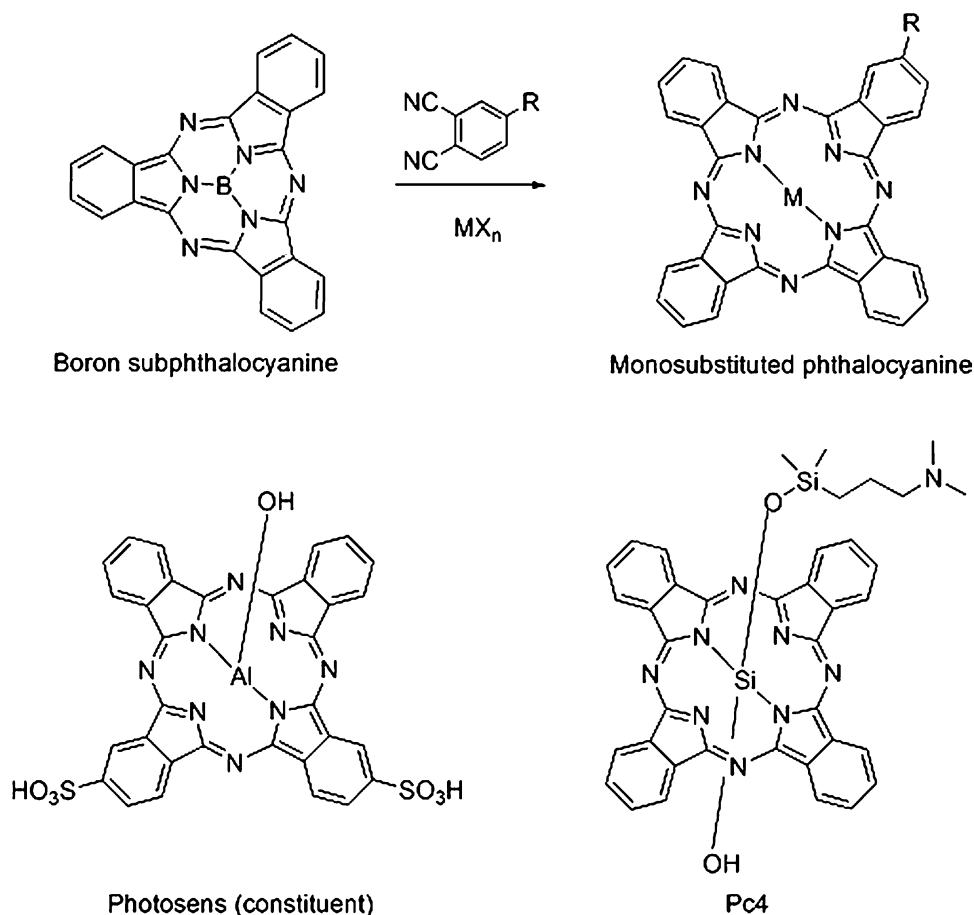
mTHPC *meso*-tetra(*m*-hydroxyphenyl)chlorin; *ALA* 5-aminolaevulinic acid, a precursor of protoporphyrin IX (PPIX)

metabolism exhibited by tumor cells means that reasonable ratios of the resulting photosensitizer are possible between normal and tumor cell populations. Local application of ALA-containing creams and optimized incubation times have produced excellent therapeutic outcomes in several forms of skin and soft tissue malignancy. As might be imagined, local application of photosensitizers in PDT does not result in the generalized photosensitivity reported with systemic administration as mentioned above.

Apart from porphyrin-type photosensitizers, the main other chemical class employed for clinical PDT is based on the phthalocyanine group. As may be seen from Fig. 3, this class of photosensitizer is structurally related to the porphyrins covered previously, but the phthalocyanines have no naturally occurring lead compounds and are an entirely synthetic class of photosensitizer. Originally discovered during industrial aromatic dye

synthesis in the 1930s, phthalocyanines were accidentally produced via the interaction of phthalic acid, ammonia, and copper from the reaction vessels, copper phthalocyanine being the first example of this type of pigment.

Phthalocyanine derivatives offer a considerable advantage over many previously used porphyrins in exhibiting significantly increased intensity of light absorption coupled with extension of absorption into both the red and near-infrared regions of the visible spectrum. Although the dye chemistry required for phthalocyanine production is well established, this is based industrially on the original copper phthalocyanine, which is an extremely stable blue pigment – thermally, chemically, and photochemically. The characteristically high photochemical stability of this class of photosensitizer is attributable to the presence of the central copper(II) ion, which has a paramagnetic electronic configuration. This



Dye: Photodynamic Therapy, Fig. 3 Phthalocyanine derivatives for PDT

provides a rapid deactivation route for the photo-excited singlet state (see Fig. 1), and consequently, insignificant amounts of reactive oxygen are formed which might otherwise decolorize the pigment. Phthalocyanine research for PDT purposes has therefore been concentrated on derivatives having diamagnetic central metal ions such as zinc or aluminum.

In most cases, phthalocyanines for colorant end use are produced as pigments. Since solution formulations are the norm in clinical use, the inclusion of suitable solubilizing chemical groups has been a significant area of development for PDT-useful derivatives. A small amount of liposomal formulation has been also used for tumor delivery of the non-functionalized material.

To produce readily aqueous-soluble phthalocyanine dyes, metallophthalocyanines may be directly sulfonated. Conversely, phthalic acid derivatives bearing water-solubilizing groups (e.g., sulfonic or carboxylic acids) may be employed. In both of these approaches, the derivatives produced are mixtures by degree of functionality and isomeric constituency. While this is acceptable in commercial dyes, it is not in pharmaceutical development – significantly, a major argument against the continued use of hematoporphyrin derivatives remains their lack of chemical purity. However, a sulfonated mixture based on chloroaluminum phthalocyanine (*Photosens*, Fig. 3) has been in use in the former Soviet Union.

The problem of isomeric functionality has been overcome via the use of subphthalocyanine chemistry (Fig. 3), allowing the introduction of mono-functionality in the ring system, and via the use of a reactive central atom such as silica, which may be functionalized both above and below the plane of the ring (e.g., Pc4, Fig. 3).

Necrosis/Apoptosis

The majority of photosensitizers employed in PDT are either negatively charged or neutral molecules which are generally oil soluble to a significant degree (lipophilic). In many cases, this means that they are extensively bound to blood lipoproteins and, consequently, concentrate in the tumor vasculature. Illumination of the tumor site at this stage thus leads to vascular damage and occlusion, starving the tumor of oxygen.

Direct damage to the tumor itself requires uptake into tumor cells. Illumination then either causes cell death by general, nonspecific oxidation, leading to tumor necrosis, or by programmed cell death (apoptosis). The latter route is more complex and involves an initial damaging event, e.g., to the lysosome or mitochondrion which then triggers intracellular messengers such as the caspases and leads to a cascade of internal processes and cell death.

Treatable Cancers

Initially, the main – and logical – target for PDT was skin cancer, given the ease of application of both photosensitizer and light to such presentations. However, the adoption of fiber optics allowed a rapid expansion of the approach into deeper-seated disease. Similarly, advances in photosensitizer design produced derivatives, as noted above, with extended light absorption characteristics. These provided a more efficient use of light since the longer wavelengths required (≥ 660 nm) lie outside the absorption spectra of endogenous absorbers such as heme and melanin. Table 1 provides an overview of the range of treatment available for malignancy and premalignancy.

Dye: Photodynamic Therapy, Table 1 Photosensitizers clinically trialed in oncology

Photosensitizer	Malignancy indication
Hematoporphyrin derivative (HpD)	Barrett's esophagus
	Bladder ^a
	Cervical ^a
	Gastric (superficial) ^a
	Lung ^a
	Esophageal ^a
	Pancreas
5-Aminolaevulinic acid (ALA)	Peritoneum
	Actinic keratoses ^a
	Barrett's esophagus
	Basal cell ^a
	Bowen's disease
<i>meso</i> -Tetra(m-hydroxyphenyl) chlorin (mTHPC)	Gynecological
	Barrett's esophagus
	Brain
	Head and neck ^a
	Lung
	Pancreas
	Prostate

^aClinically approved

Photodynamic Antimicrobial Chemotherapy (PACT)

The negative effect of illumination on a stained microorganism was first demonstrated by Raab in 1900, thus predating PDT [3]. However, while porphyrin-based PDT has become a significant clinical field, the application to microbial pathogens has taken over a century even to become a clinical reality on a small scale. The main reason for this is the straightforward access to effective “conventional” antimicrobial agents enjoyed by the affluent countries of the world over the past 60 years. Only in the twenty-first century, with the waning efficacy of many drugs, is there a requirement for alternative approaches to infection control, particularly in view of bacterial resistance to conventional drugs [4], such as penicillins and tetracyclines.

Logically, as PACT uses the same approach as PDT but applied to microbial pathogens, its limitations are similar also. Thus, the photosensitizers employed in or proposed for PACT, known collectively as “photoantimicrobials” due to their extremely broad spectrum of activity, require local or topical application due to the light activation factor. While this might, at first glance, appear to trivialize the approach, as it cannot thus be used for systemic infection, this is not the case. Following the PDT paradigm, localized infection may be targeted anywhere in the body via the use of endoscopy and fiber optics. Indeed, initial forays into the clinic have been made in oral and lung infection, as well as the more obvious dermatological areas, such as acne, ulcers, and nail infection.

Photoantimicrobials

One area of dissimilarity between PACT and PDT lies in the chemical types involved. While PDT has remained firmly based on the porphyrin class from which it was first developed, photosensitizers for use against microbial species require different makeup for broad-spectrum activity. However, it should be noted that, since the metabolism of many bacterial species includes porphyrin synthesis, ALA may be of use here also, used in the same way as detailed above for anticancer PDT.

The use of dyes as chemotherapeutic agents in the first half of the twenty-first century was a result of scientific development of the heavily dye-oriented organic chemistry begun by Perkin, Hoffmann, and others. The first dye to be used clinically was the phenothiazine *methylene blue* (C.I. Basic Blue 9), which was shown to be antimalarial in 1890. The chemical structures of many modern antimalarials, as well as drugs for the central nervous system, have much in common with methylene blue as a consequence, the phenothiazine dye being tested as a treatment for many different illnesses mainly due to its very low toxicity in humans. Methylene blue has also been the lead compound in photoantimicrobial research since it was shown to be active against bacteria, viruses, fungi, and protozoa by various

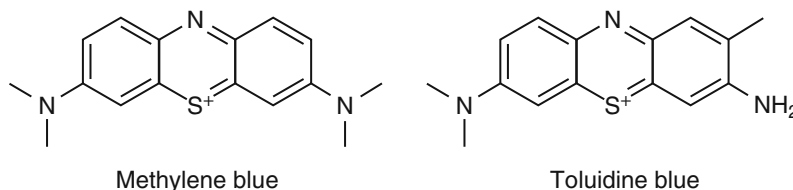
groups before the mid-1930s. Unfortunately for the field of photoantimicrobial endeavor, 1935 saw the initial clinical release of the first conventional antibacterial agent, sulfanilamide (as *prontosil*) by IG Farben, beginning what has been known to many as the “antibiotic era.” Clearly, effective and easily administered antibacterial agents are more appealing than an approach requiring light activation, and photoantimicrobial research languished as a result.

Unsurprisingly, given the flagrant overuse of our valuable antimicrobial arsenal, drug resistance – particularly in bacteria – has become a very serious clinical problem. Indeed, the problem is increasing to such an extent that many strains of bacteria in our hospitals are approaching indestructibility, as far as conventional antibacterial drugs are concerned. This resistance is merely the outcome of evolution directed by overuse, and misuse, of drugs which have only a single site/mode of action, i.e., which may be circumvented by a single alteration as a result of natural bacterial mutation. One of the advantages of the photoantimicrobial approach is that there are multiple sites and modes of action, making microbial resistance development far less likely.

Broad spectrum, in terms of photoantimicrobial agents, means activity against bacteria, viruses, fungi, and protozoa, and this is generally associated with positively charged (cationic) photosensitizers, such as methylene blue and its derivatives. One of the reasons that the porphyrin photosensitizers utilized in anticancer PDT are not useful leads for photoantimicrobial development is that, as anions, they have little effect on gram-negative bacteria and so cannot be considered to be broad-spectrum agents. However, even this level of activity is massively different to that of conventional antimicrobial drugs, which are *either* antibacterial *or* antiviral *or* antifungal *or* antiprotozoal.

Methylene blue itself is a tricyclic heteroaromatic molecule which absorbs red light in the region of 660 nm. The related dye *toluidine blue* (C.I. Basic Blue 17) absorbs at a shorter wavelength (c. 625 nm), mainly due to the non-alkylation of one of its amino groups (Fig. 4). Both of these photosensitizers produce significant yields of singlet oxygen, both in vitro and in vivo.

Dye: Photodynamic Therapy,
Fig. 4 Phenothiazinium photoantimicrobials



While there have been considerable strides made in synthesizing improved agents with these dyes as lead compounds, the amount of regulatory testing required has so far hindered progress toward the clinic, and methylene blue appears to be the chosen derivative, having been licensed for oral disinfection and blood plasma decontamination in several countries. The extension of such applications to bacterial decolonization and other local anti-infection measures is proposed in order to lessen the burden on conventional antimicrobial delivery, particularly in view of the resistance problem and the dwindling number of agents effective against serious diseases such as septicemia and bacterial meningitis.

Light Sources

Early clinical PDT was carried out normally using in situ hospital lasers, which were expensive both to purchase and to maintain. While the advent of much cheaper diode lasers improved the situation financially, both sources were tied to single-wavelength output and were thus restricted in the number of photosensitizers which might be used. However, both tunable dye lasers and light-emitting diode (LED) sources have revolutionized the area, in terms of adaptability and cost, such that financial considerations present no bar to the adoption of the technique by healthcare concerns. It should be noted that there are still only three photosensitizers generally licensed for clinical use (Table 1).

Treatment of Age-Related Macular Degeneration (ARMD)

Photosensitizers can be used in the treatment of wet, age-related macular degeneration – a condition, normally in the older patient, where new

blood vessels (neovasculature) grow from the back of the eye and break through into the region of the retina containing the rods and cones necessary for the perception of color and shading. The new vessels leak blood and fluid, causing scarring which is associated with a gradual loss of central, but not peripheral, vision.

The concentration of lipophilic, anionic (typically porphyrin-type) photosensitizers in tumor vasculature has been mentioned. This propensity has a positive spin-off in the treatment of wet, age-related macular degeneration. The use of lipophilic, anionic photosensitizers such as chlorin *e*₆ (verteporfin) allows selective targeting and ultimate destruction of the neovasculature.

Cross-References

► [Dye, Functional](#)

References

1. Brown, S.B., Brown, E.A., Walker, I.: The present and future role of photodynamic therapy in cancer treatment. *Lancet Oncol.* **5**, 497–508 (2004)
2. dos Santos, A.F., Queiroz de Almeida, D.R., Terra, L.F., Baptista, M.S., Labriola, L.: Photodynamic therapy in cancer treatment - an update review. *J. Cancer Metast. Treat.* **5**, 25 (2019)
3. Wainwright, M.: Photodynamic antimicrobial chemotherapy (PACT). *J. Antimicrob. Chemother.* **42**, 13–28 (1998)
4. Wainwright, M., Maisch, T., Nonell, S., Plaetzer, K., Almeida, A., Tegos, G.P., Hamblin, M.R.: Photoantimicrobials – Are we afraid of the light? *Lancet Infect. Dis.* **17**, e49–e55 (2017)

Dyeing

► [Coloration, Textile](#)

Dyes

- [Inorganic, Hybrid, and Functional Pigments](#)
-

Dye-Sensitized Solar Cells

- [Dye, DSSC](#)
-

Dynamic Color Design

- [Environmental Color Design](#)
-

Dynamic Color Spreading

- [Motion and Color Cognition](#)
-

Dynamics of Color Category Formation and Boundaries

Stephanie Huetter
Department of Psychology, University of
Memphis, Memphis, TN, USA

Definition

Dynamics of color boundaries is broadly the area that characterizes color categorization as a process that occurs over time.

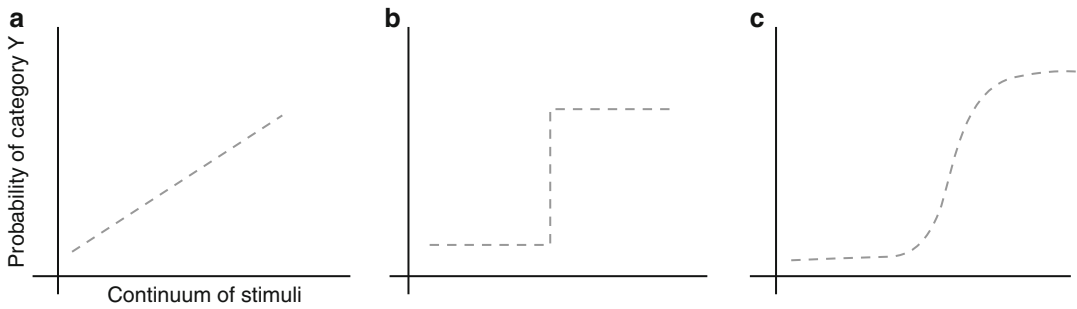
Color Boundaries

A color category boundary is the area in perceptual color space where the labeling of one color transitions into another label. Color is a continuous wavelength dimension but distorted into non-linearly perceived perceptual categories. A category boundary is fundamentally a distortion

of continuous color space during perceptual and cognitive processing of color. There is no external signal in the wavelength dimension to tell the perceptual system to make this abrupt transition. This is called *the problem of invariance* [1]. The demarcation of color space has led to empirical work seeking to characterize color boundaries and discover their underlying dynamics.

There are various tasks that can be used to define a color boundary. One way researchers can do this is simply by presenting a color swatch and asking participants for the appropriate color label. Boundaries can also be measured by presenting two color swatches and asking whether they are the same or different, which is called *discrimination* (see ► [“Color Categorical Perception”](#) for more details). The categorization graphs show a curve that can be classified as one of three types: step, sigmoid, or linear (Fig. 1). The step function is a theoretical ideal not found in empirical data. The discrepancy between the theoretical ideal and obtained data may be due to some meaningless noise in the perceptual system and appears in the data as a sigmoid, even though the actual underlying transition may be a discrete transition. Alternatively, the sigmoid may accurately reflect the boundary: The slope of the sigmoid is indicative of how gradual the transition is between categories. Finally, there could be a linear perception of the stimuli, which would indicate not having categories.

Human experimental data displays sigmoidal categorization boundaries, where stimuli near the boundary are not always reliably categorized into the same category. This area near the boundary is unique, because it represents a place in perceptual space where there is a distortion. This makes between-category distinctions a larger perceptual distance apart, effectively compressing within-category distances between stimuli and expanding between-category distinctions. One account states that categorical distortions are independent from and occur in parallel to continuous hue processing [2]. This issue bears upon whether categorical distortions have an effect on the perception of color or if they are a cognitive phenomenon induced by labeling rather than a warping of perceptual space [3].



D

Dynamics of Color Category Formation and Boundaries, Fig. 1 Panel (a) illustrates a linear or identity function, panel (b) the step function, and panel (c) the sigmoid. A linear function represents sensitivity to continuity and a lack of categorization. The step function is the most extreme categorization, where there is an

instantaneous transition from one category into another category. The sigmoid in panel (c) represents the more gradual transition from one category into another. In each panel, the *x-axis* represents a continuum ranging from category X to category Y, and the *y-axis* represents the probability or activation value for category Y

Further support for the role of labels in the formation of category boundaries comes from comparative work on color perception. Color category boundaries do not arise from pre-cortical processing, as there is no invariant signal coming from opponent process cells or the spectral sensitivities of cones themselves. Thus, it is unlikely that boundaries arise from early low-level visual processes likely to be found in other primates. Baboons do not have color boundaries; rather, they perceive hue linearly, even though their retinal and opponent cell processing are similar to humans ([4]). This evidence supports the role of higher level cortical processing and lends support to the Whorfian hypothesis (also called linguistic relativity). There is also evidence that color boundaries are learned and not innate further emphasizing a strong role for labels [5].

A dynamic view of color boundary formation is well supported by many already existing theories, including linguistic relativity for more detail). This hypothesis acknowledges that the influence of language labels exerts some influence over the perception and formation of boundaries [6]. This influence could modify a boundary in many ways: Two labels could sharpen the boundary between them, making the sigmoidal slope steeper and expanding perceptual distances across the category boundary. One label could make the slope shallower or even linearly perceived (non-categorical). Finally, a label could translate

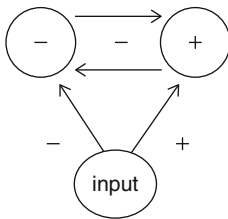
the boundary, acting as an attractor or repeller for the boundary.

These kinds of dynamics are easily conceptualized as linguistic. However, many other factors could exert an influence on boundaries, such as biology, environment, or context. These alterations of the boundary could also happen at various timescales, on the order of minutes (adaptation; [7]), years (lifespan; [8]), or much longer spans (evolution; [9]). As opposed to thinking of color as something static and invariant, a dynamic view instead emphasizes the role of change and how those occur over time. The typical measurement of boundaries does not reflect influences of context that can alter the boundary [10].

Within a category hues are labeled the same. This could be because they are processed equivalently or because the task forces participants to place the stimuli into one of two categories. If they are processed equivalently, the underlying activation of the category should be all or none: For example, on a given trial, a given hue would be processed as entirely “green.” Alternatively, both color categories may be active to various degrees. For example, the hue would be processed probabilistically as 0.8 “green” and 0.2 “blue.” Over hundreds of milliseconds, the pressure of being required to put the hue in one of two categories would create competition between the two, finally resolving into a discrete answer of either “green” or “blue.” This latter process is called gradient

categorization and was shown to apply to color perception [11].

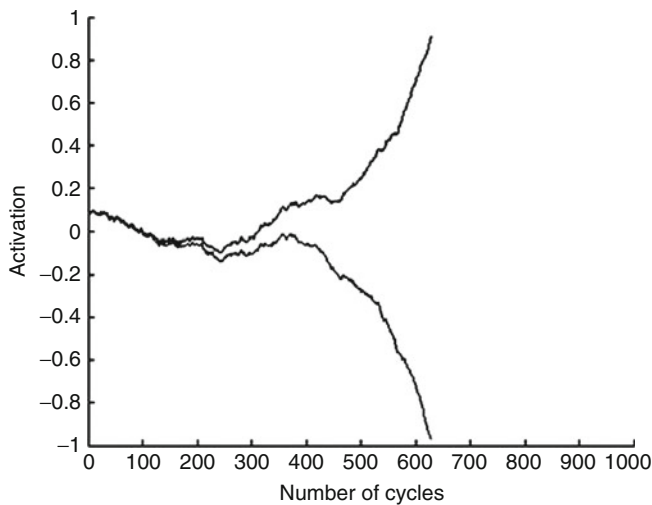
One study demonstrated gradient categorization in color perception using a computer mouse tracking paradigm [11]. A color categorization task in which participants would begin moving vertically up the screen with a computer mouse and a color from a continuum of green to blue would appear briefly on the screen. The participants would continue moving up the screen and indicate their response by moving over a prototypically blue or green swatch in one of the top



Dynamics of Color Category Formation and Boundaries, Fig. 2 The inputs range from -1 to $+1$ which represent a range of hues to be categorized. The second layer represents the two category options, and the lateral connections between them perform the competition

corners. The trajectories of computer mouse movements show subtle deviations toward the competing category response option – for example, veering toward green when the final response is blue. These deviations reflect the maintenance of continuous hue detail used to help make a discrete, categorical response. The categorical end goal and the gradation in processing coexist in the competition model and parallel human performance.

To demonstrate the maintenance of some form of continuous detail, whether that is in terms of raw hue value or in terms of distance to the category boundary, there must be evidence of a monotonic relation to the category boundary, as measured by some behavioral channel. This kind of signal is not seen in measurements taken only once per trial, as in typical naming tasks. A category boundary of some sort is mathematically guaranteed in a task requiring a decision between one of two items: Thus, within-category subtleties in processing are often neglected by measuring the boundary, which is informative about what is happening near the boundary (e.g., the steepness of the slope indicating the amount of distortion near the boundary).



Dynamics of Color Category Formation and Boundaries, Fig. 3 Competition over time: the two nodes activations are plotted over a number of cycles. You can see the activations begin very similarly, and then gradually one

begins to win, and it rapidly takes the activation away from the other node. This simulation contains a random noise term, which is responsible for the smaller fluctuations in each line, and can produce variability in the final response

Graded categorization can be computationally modeled as competition or even with an evolutionary algorithm emphasizing the role of motor movements [12]. Competition can be implemented with lateral inhibition, meaning that two representations are connected laterally and mutually inhibit one another (Fig. 2). This kind of a network shows gradations in performance during perceptual choices [13]. This network will cycle through the lateral connections, gradually taking the category with the highest activation and increasing its activation by stealing from the other node (Fig. 3). A mechanism like this is hypothesized to underlie the formation of discrete responses, and these gradations in initial activation can be seen in motor responses that are sampled at a very high density [11]. The competition model serves as a mathematical proof that one could accomplish a task such as categorical perception with minimally complex circuitry.

Category boundary formation is a dynamic phenomenon susceptible to context such as labels, objects, and the task being used. The typical color boundary is a relatively steep sigmoidal shape, and the steepness of this slope indicates the degree to which the perceptual or cognitive system has warped the continuous hue space. This could be due to competing labels or prototypes. The dynamic view emphasizes the characterization of these boundaries over many different timescales, and many different areas of color perception address these kinds of dynamic influences (Linguistic relativity, top-down influences, evolution of color categories, etc.) Future directions in this area may look more closely at these variables, and seek to characterize the process in real time, using dense-sampling methods appropriate to the timescale of the phenomenon.

References

1. Harnad, S. (ed.): *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, New York (1987)
2. Bornstein, M.H., Korda, N.O.: Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychol. Res.* **46**, 207–222 (1984)
3. Pilling, M., Wiggett, A., Özgen, E., Davies, I.R.L.: Is color “categorical perception” really perceptual? *Mem. Cogn.* **31**, 538–551 (2003)
4. Fagot, J., Goldstein, J., Davidoff, J., Pickering, A.: Cross-species differences in color categorization. *Psychon. Bull. Rev.* **13**, 275–280 (2006)
5. Özgen, E., Davies, I.R.L.: Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol. Gen.* **131**, 477–493 (2002)
6. Roberson, D., Davidoff, J., Davies, I.R.L., Shapiro, L.R.: The development of color categories in two languages: a longitudinal study. *J. Exp. Psychol. Gen.* **133**, 554–571 (2004)
7. Bornstein, M.H., Korda, N.O.: Identification and adaptation of hue: parallels in the operation of mechanisms that underlie categorical perception in vision and audition. *Psychol. Res.* **47**, 1–17 (1985)
8. Raskin, L.H., Maital, S., Bornstein, M.H.: Perceptual categorization of color: a life-span study. *Psychol. Res.* **45**, 135–145 (1983)
9. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)
10. Goldstone, R.L.: Effects of categorization on color perception. *Psychol. Sci.* **6**, 298–304 (1995)
11. Huette, S., McMurray, B.: Continuous dynamics of color categorization. *Psychon. Bull. Rev.* **17**, 348–354 (2010)
12. Beer, R.D.: The dynamics of active categorical perception in an evolved model agent. *Adapt. Behav.* **11**, 209–243 (2003)
13. Usher, M., McClelland, J.L.: The time course of perceptual choice: the leaky competitive accumulator model. *Psychol. Rev.* **108**, 550–592 (2001)

Cross-References

► [Color Categorical Perception](#)

Dyschromatopsia

► [Color Perception and Environmentally Based Impairments](#)

E

Early Color Lexicons

► [Ancient Color Terminology](#)

Early Expression of Color in Language

► [Ancient Color Terminology](#)

Eco-/genotoxicity of Synthetic Dyes

► [Colorant, Environmental Aspects](#)

Effects of Color Terms on Color Perception and Cognition

Jonathan Winawer¹ and Nathan Witthoft^{1,2}

¹Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

²Department of Psychology, Stanford University, Stanford, CA, USA

Synonyms

[Categorical perception](#); [Linguistic influences on color perception](#); [Sapir-Whorf hypothesis](#)

Definition

The possibility that naming colors, either in a single instance or habitually over a lifetime, alters color perception.

Color Perception and Color Communication

When we communicate about the colors of scenes and objects comprising our visual experience, what we see informs our choice of words. A question that has interested many cognitive psychologists is whether the color words we use affect how we see. One view is that naming is for communication and has no effect on how we see or experience the world. In this view, color appearance, and therefore performance on tasks which strictly depend on color appearance, is determined entirely by the visual system and not at all by the language that one speaks. Color vision is informationally encapsulated, its output is automatically produced, and while its output is available for cognitive processes like decision making and speaking, higher cognitive processes cannot alter its process [1]. This can be contrasted with the idea that the very act of naming a color, either in a single instance or habitually over a lifetime, can change one's perceptual experience with the result that colors assigned the same label (i.e., belonging to the same color category) are more difficult to discriminate than colors assigned

different labels (belonging to different color categories).

The two views have persisted in the psychology literature for at least two reasons. First, the empirical evidence has been mixed, with some researchers finding effects of color terms on perceptual or cognitive tasks and others failing to find such effects. Second, when positive evidence of the influence of color naming on color perception has been found, it has often been unclear at what level the effects are manifested. For example, the same experimental results might be interpreted by some researchers as evidence that color terms affect how one sees colors and by other researchers as evidence that color terms affect how one remembers or makes judgments about colors. In this entry, we examine the question of how to design experiments to probe the effects of color terms on color perception and cognition and how to interpret the effects.

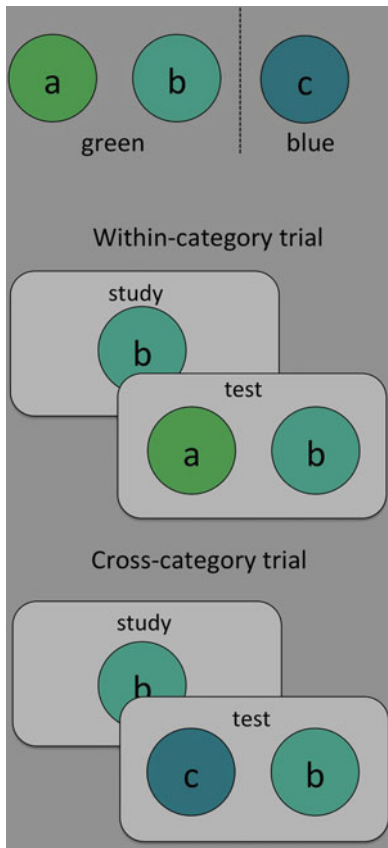
The study of categorical effects in color perception has its roots in two historical traditions. The first is the Sapir-Whorf Hypothesis, a concept that originated in anthropology [2]. Whorf and Sapir formulated this hypothesis in different ways in their writing, but researchers have generally taken their view to be something like the way in which experiences are mapped onto words affects the experiences themselves, so that speakers of different languages perceive and conceive of the world differently. A second and related historical tradition is the study of categorical perception [3, 4] and in particular the study of phoneme perception in language. Distinctions between certain phonemes, such as /r/ and /l/, are necessary for speaking some languages, such as English, but not other languages, such as Japanese. Adult English speakers are better at the /r/ versus /l/ discrimination than Japanese speakers (though Japanese speakers are above chance, meaning that the effects are not completely categorical). In phoneme learning, it appears that the learning is primarily a loss rather than a gain; the /r/ and /l/ sounds become more similar as a child learns Japanese rather than more different as a child learns English [5]. Psychologists studying color have similarly asked whether making

color distinctions improves sensitivity to distinguish shades that straddle a category border or worsens one's sensitivity to distinguish shades that fall within a single color category.

No color scientist takes seriously the idea that color perception is as categorical as color naming. It is trivial to find two colors that are both called green that can be distinguished from one another. A more nuanced question is whether colors that are named alike show any degree at all of increased perceptual similarity (or decreased discriminability) compared to colors that are named differently.

Breaking the Circularity: The Logic of Measuring Category Effects

To assess the effects of color categories on perception and cognition, one has to sort out the direction of causation: a pair of colors may belong to different categories because they look very different, rather than looking different because one learns to put them in different categories. In fact, it is certain that this happens; colors that look alike are more likely to belong to the same color category, and language groups across the world tend to assign colors to categories in similar ways, though some differences exist [6]. To know whether there is *also* an effect in the reverse direction, that is, an effect of category on perception or cognition, one needs to have a prediction of how similar the colors should be in the absence of a category effect. Given such a prediction, one could ask the following: are two colors from the same category (Fig. 1a, b) less discriminable (or more similar) than two colors from different categories (Fig. 1b, c), even when the two pairs are identically spaced? That is, are “within-category” trials harder than “between-category” trials? This leads to a potentially circular problem. If the stimulus pairs are chosen such that they are equally spaced, then are they not guaranteed to be equally similar? Or, conversely, if one pair is shown to be more discriminable than the other, does this mean that the color space which placed them at equal distances from one another was flawed?



Effects of Color Terms on Color Perception and Cognition, Fig. 1 Testing effects of color terms on a color task. The *upper* panel shows three stimuli that might be used in an experiment to investigate whether color terms affect a particular task. If the task involves memory, the subject may be instructed to first remember a color (“study”) and then identify the remembered color (“test”). If the incorrect stimulus comes from the same category, the trial is considered “within-category.” If the incorrect stimulus comes from a different category, it is considered a “cross-category” trial. The colors were chosen so that *a* and *b* look *green* and *c* looks *blue*, but the observer may not agree depending on the display and individual differences between observers

This entry summarizes the answer to two questions pertaining to whether and how color naming affects color cognition:

1. How are experiments designed to break the problem of circularity posed above?
2. If the problem of circularity can be broken, then what aspects of color perception and cognition show effects of color names?

Model-Based Approaches

One way to break the circularity problem in studying the effect of color names on color cognition and perception is to compare the results to a null model, meaning a model which does not use color categories as part of its machinery even as it tries to predict them. The choice of color space for a null model is an essential consideration. If the space is designed to reflect perceptual similarity, such as the Munsell color space, then it may already incorporate category effects. And if such a space fails to predict performance on a color judgment task, then it may simply indicate that the space did not perfectly achieve its design goal, rather than something about color categories. Hence, a null model is of limited value for the study of effects of language on color processing if the model is purposely shaped to account for the pattern of human behavior, such as the Munsell system or the CIELab space.

In contrast, a model that sticks more closely to peripheral aspects of color processing can be informative. For example, Brown and colleagues used the MacLeod-Boynton space to predict reaction time in a visual search experiment for an oddball color [7]. The MacLeod-Boynton space represents color stimuli as a linear transform of cone excitations. It contains no information about color names or categories or other properties of color cognition. In the visual search experiments, one color (the target) differed from the other colors (the distractors), and the time to find the target was modeled as a function of the difference between the target and distractors in MacLeod-Boynton space. Trials in which there was a large difference between the targets and distractors in the space were predicted to have a faster reaction time. Some trials were cross-category and some within-category, analogous to the design in Fig. 1. The authors found that the null model provided a good fit to the empirical results, with no systematic deviations for cross-category trials. For this color task, color discrimination time could be explained by a model based only on cone absorptions, and hence there was no evidence that discrimination was affected by color categories. The null model succeeded.

It is important to consider what happens when a null model fails. Suppose, for example, a model predicts that stimuli a and b and stimuli b and c in Fig. 1 are equally discriminable, and yet experiments show that cross-category pairs are discriminated faster or recalled more accurately than within-category pairs. While additional mechanisms or a different model may fit the data better, it does not follow that color categories made the stimuli more discriminable. It may simply be that the model was incorrect or inappropriate for the task.

Equating Color Stimuli but Varying the Subject Population

Another way to break the circularity problem in studying the effect of color names on color cognition and perception is to test identical sets of color stimuli with speakers of languages with differing color terms.

For example, for English speakers there is no common color word that refers to all three stimuli in Fig. 1. Speakers of some languages (sometimes called “grue” languages), however, could refer to all three colors by the same name. An experimenter can therefore avoid the difficulty of perfectly equating differences between the stimulus pairs: for speakers of languages like English that place a boundary between stimuli b and c, the b/c pair should be more different than the a/b pair *relative to* speakers of “grue” languages. The colors are controlled not by the exact localization of colors in a color space but rather by showing exactly the same sets of colors to different groups of observers and seeing if their pattern of responses differs. One of the challenges in such experiments is finding appropriate subject populations. Color naming patterns are quite similar across many languages. Nonetheless, large differences in naming patterns can be found by comparing populations from an industrialized western society such as the UK or the USA with speakers from more remote cultures which have had minimal contact with western societies, such as the Dani and Berinmo of Papua New Guinea or the Himba of Namibia. However, even among modern industrial societies, differences can exist. For example, Russian speakers use different

words for the colors light blue (“goluboy”) and dark blue (“sinij”), while English speakers may refer to both as “blue.”

The results of cross-linguistic studies have not been uniform. Some have found effects of language on a color task, such as a color memory task comparing English speakers with Berinmo speakers [8]. Others have found little to no effect. One reason is that the effects, even when found, tend to be small compared to the shared properties of color processing found in all observers. For example, consider the effect of metamers: two colors with quite different spectral power distributions will be indistinguishable to an observer if the two colors result in the same pattern of cone excitations. This is a very large effect and is quite consistent across observers. A metameric pair for one person is likely to be a metameric pair for another person, or nearly so, unless one of the observers has a color vision deficit. Put another way, three-channel color displays such as television do not need to be tailor-made according to the color naming patterns within a language community. Color metamers do not differ with language because the information lost at the level of cone absorptions cannot be regained. The cross-linguistic differences that have been observed tend to be quite modest in magnitude relative to the much larger effects shared across all observers.

Equating Color Stimuli but Varying the Hemifield: Lateralized Whorf

Rather than testing for an effect of language by holding the stimuli constant and varying the language of the subjects, some experimenters have compared the same speakers but on different sides of visual space. In these experiments, ostensibly the same set of color stimuli are shown on either the left or the right side of a fixation point. The logic of this kind of experiment is that verbal processes are computed primarily in the brain’s left hemisphere, and visual inputs are routed to the contralateral hemisphere, so that stimuli to the left of fixation are processed more in the right hemisphere, and vice versa. The separation by hemisphere is not complete for either vision or language; nevertheless, researchers hypothesized

that if effects of color terms on color tasks were to be found, they would be larger for stimuli presented to the right of fixation than the left of fixation. A number of reports have supported this hypothesis, known as the lateralized Whorf effect [9]. Other reports have failed to confirm the finding, however, and the reasons for the discrepancies have not yet been resolved [7, 10].

Equating Color Stimuli but Varying the Role of Language: Verbal Interference

Another strategy to break the circularity of measuring a category effect is to use the same set of stimuli under different conditions designed to make it easier or harder to use language in the task. For example, a color task can be conducted under conditions in which the subjects' verbal faculties are engaged with a secondary task, such as reading a list of words aloud ("verbal interference"). If a secondary, language-related task reduces a putative category effect, then the problem of circularity is solved for this experiment. There is no need to precisely equate the distances between color pairs; rather, the role of language is implicated because performance changes when the availability of language changes. Experiments have shown that manipulating the availability of language during a color task can reduce the effect of color terms on the task.

The classic experiment of this type was done by Kay and Kempton [11], who compared English speakers to Tarahumara speakers from northern Mexico. The Tarahumara use the word "siydney" to name both green and blue. Subjects from both groups were shown triads of color chips which straddled the blue/green boundary and asked to select the chip which was least similar to the other two. They found that English speakers tended to exaggerate the distance between chips which straddled the boundary, while the Tarahumara did not. This finding is consistent with the idea that the availability of the terms "blue" and "green" influenced the English speakers' decisions about which color chip was the most dissimilar when one of the chips came from a different category. Kay and Kempton proposed that rather than resulting from some difference in color appearance, the English speakers were

relying on the linguistic distinction to make decisions when the perceptual information was insufficient. To test this idea, they repeated the task, but only allowed the subjects to compare two of the chips at a time. The pairs were set up so that the chip in the middle of the color range could be alternately compared with one of the two to either side. This meant that if the middle chip was at a boundary, it would seem bluer than one chip and greener than the other, thus eliminating the usefulness of having a verbal code. Under these conditions, the English speakers showed the same performance as the Tarahumara.

If a category effect goes away when labels become unavailable or not useful, then it is unlikely that the effect is due to color terms affecting early perceptual processes. While such an account is logically possible, it would require color appearance to be altered only during those moments when one is accessing the labels. A more parsimonious explanation is that the *decision* process is affected by language. Verbal labels may be used to help keep track of the various stimuli in an experiment, either over a memory delay or when comparing stimuli spread over space. If, on a particular trial, all the stimuli come from the same verbal category (e.g., they are all blue), then labels are unlikely to help accomplish the task (and might even hinder performance). In contrast, if stimuli in a trial can easily be assigned with different labels (e.g., one blue and one green), then access to the labels may facilitate memory or the comparison process. If a verbal dual task interferes with the ability to label stimuli, even implicitly, then this may eliminate one strategy or source of information for accomplishing the task and hence may change performance. Thus, verbal interference effects are more likely to reflect a role of color terms on decisions, strategy, and memory, rather than perception.

Cognition and Perception: What to Measure

Thus far we have primarily been focused on how to break the problem of circularity when measuring the effect of color terms on color tasks. Here we focus on the tasks themselves. What kinds of

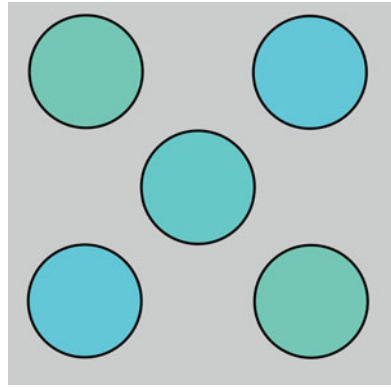
tasks are used to test for the effects of color terms, and what do the results show?

Subjective Judgments

Perhaps the most natural way to ask whether color terms affect color perception is to ask subjects to make direct, overt judgments about color stimuli. For example, one could present subjects with a variety of color samples and ask them to place into groups those that are most similar. Experiments like these have been conducted with observers across languages. Grouping patterns tend to be similar across language groups, but some differences are found consistent with grouping together colors that are given the same name in the language. The difficulty with this type of result is that subjects may choose to group together colors because they have the same name, rather than because they look most alike. Hence, it is difficult to identify whether the effect of color terms is on perception, or on the strategy of how to form groups.

Perceptual Grouping. A more indirect approach can make it less likely that subjects use a naming strategy to make a decision. For example, Webster and Kay took advantage of the tendency of human observers to perceptually group together objects that have a similar color [12]. By arranging disks into two diagonals with an ambiguous central disk, participants could indirectly report perceptual similarity by indicating whether the disks grouped into a line sloping up to the right or up to the left (Fig. 2). The question they then asked was whether the decision to group dots into a triplet was influenced by color categories, specifically whether there was a bias to group the dot with the diagonals when the three came from the same color category, regardless of the proximity in color space. These effects were compared to a null model based on a linear transform of cone excitations, similar to the approach used by Brown and colleagues [7]. The null model did a good job explaining the data, and hence there was no evidence that color categories modulated the task.

Hue Scaling. Interestingly, a different measure of color perception using the same null model did show a significant deviation from the model. This measure is called “hue scaling,” in which subjects view a color sample and rate the perceived amount



Effects of Color Terms on Color Perception and Cognition, Fig. 2 Gestalt grouping configuration used to study the effect of color categories on color appearance [12]

of red, green, blue, and yellow in a stimulus. Compared to a null model, colors that fell on the blue side of the blue/green border were effectively rated as containing more blue than would be expected purely from the model. That is, the results were biased away from the null model in the direction predicted by an effect of color terms on the task.

Subjective judgments are a desirable way to measure perception because they are direct: if one wants to know what something looks like to a subject, then asking for a subjective judgment is a sensible way to find out. On the other hand, because the judgments are subjective (there is no right answer) and there is no time pressure, these tasks are also amenable to strategic decisions that have little to do with perception. For example, when subjects are explicitly asked to rate the amount of blue in a stimulus, the internal mapping from perception to a judgment may be influenced by the fact that the subject knows that the stimulus falls in the category “blue.”

Memory

Memory tasks designed to probe the effect of color terms on perception or cognition often take the form of remembering a color over a delay and then trying to choose between the same stimulus and a second stimulus, where on some trials the second stimulus comes from the same category and sometimes a different category. Memory tasks have sometimes been used in cross-linguistic

studies and sometimes in studies that included dual tasks (verbal interference). Several memory studies have found effects of color category on performance (e.g., [8, 13]). These studies are consistent with the idea that color terms are used as a dual code in memory; if the subject remembers the term in addition to the appearance, then the subject will be more accurate when the memory test includes only one option from the expected color category.

Reaction Time on Suprathreshold Discrimination

In a suprathreshold discrimination task, subjects typically judge which color patch matches a test color. There is only one right answer, and the discrimination is usually well above threshold so that subjects are expected to make the right answer most of the time. The measurement of interest is reaction time, and the typical experimental question is whether reaction time is faster for cross-category trials than within-category trials. In order to avoid the circularity problem, the experimenter can use one of the strategies described above, such as cross-linguistic comparisons, dual tasks, or varying the visual hemifield to which the stimuli are presented. Alternatively or in combination with these strategies, the experimenter can provide a null model to predict the data.

Suprathreshold experiments have had mixed results. Some have reported an effect of color categories on performance and some have not. When an effect of color categories is found in a speeded suprathreshold discrimination task, there are typically two alternative explanations the experimenter is confronted with. The participant may be faster on cross-category trials because stimuli that come from different categories look more different or the participant may be faster on cross-category trials because a decision is easier to make when color labels support the decision. As discussed above, if a verbal dual task eliminates a category effect, then the effect probably is in the decision stage rather than in the percept.

Subjects' experience with the task may also modulate the contribution of color categories on performance. For example, Witzel and Gegenfurtner [14] found that suprathreshold

judgments were facilitated by category differences between stimuli at many name boundaries. These effects were not apparent in a second group of subjects tested on the same stimuli, but who were trained ahead of time on difficult boundary discriminations. The trained group was also faster and more accurate on the suprathreshold task. One interpretation is that categorical labels were not useful in the training phase, and therefore these subjects learned to make the decision without waiting for the support of a categorical label.

Discrimination Thresholds

Threshold discrimination experiments are among the least ambiguous experiments in psychology. If an observer can discriminate two stimuli, then we can be certain that the observer's perceptual system has encoded the two stimuli differently. If the stimuli are indistinguishable (below threshold), then information distinguishing the stimuli was either not encoded or was lost in subsequent processing. If discrimination thresholds were altered by the color terms in one's language, this would provide the most direct evidence that color terms affect perception of colors. Experiments that have compared threshold discrimination among speakers of languages that divide the color spectrum differently have found no effect of color terms [15].

Interpretation and Future Directions

We began with the question of whether the words we use to describe colors affect how we see. A major challenge in addressing this question is sorting out the direction of causality: since similar colors are likely to be given the same name, and dissimilar colors are likely to be given different names, how does one measure whether naming has any effect on how the colors appear? Several methods have been used to address this problem, including the use of null models, cross-linguistic experiments, verbal interference, and hemifield manipulations.

Despite some discrepancies in the literature, some general patterns are emerging. A variety of tasks and experiments have reported effects of

color terms on color cognition in the domains of judgment and decision making and memory. Results are consistent with the idea that color terms can serve as a dual code in conjunction with sensory representations for certain kinds of judgments. Consider the case of speeded discrimination tasks, discussed earlier. In these tasks, there are two sources of information that can inform a decision whether two colors are the same or different. One is a sensory representation and the other is a label. In some situations, for example, when reaction times are slow and the two samples belong to different categories, the labels are useful and may speed reaction times. In other situations, such as two samples having the same color label, or subjects who are trained to respond faster than the labels arise, the categories are not useful and therefore will not contribute to the decision or may even interfere. This kind of explanation – that a dual code explains many effects of color terms on task performance – implies that the effect of color terms is on the perceptual decision-making process, and not on the sensory representations.

In contrast, there is no firm evidence that color appearance or color threshold discrimination is affected by color terms. The pattern of results in this way differs from phoneme discrimination. This may reflect the fact that phonemes have a categorical function: words are combinations of phonemes, and if a phoneme is mis-categorized, then the word may be misheard. In contrast, objects and scenes are not combinations of color categories. There is generally no need to categorize or label a color in order to recognize an object or scene. The fact that we *can* categorize colors by labeling them may have implications for how we remember or reason about colors, but because labeling is not an inherent part of a visual process, perhaps we should not expect it to have a significant effect on visual appearance or discrimination.

In sum, there is a growing body of evidence that color terms affect some tasks involving color cognition, but no strong evidence that color terms affect sensory representations. Put another way, the closer a color task is to language, the more likely it is that it will be affected by the terms in one's language. Hue scaling involves

explicit judgments about color names, and so it is more likely to be affected by color terms. Memory is more likely to involve linguistic codes than perception, and so memory tasks are more likely to show effects of color terms. For any given task, it might not be immediately obvious whether linguistic codes are used. Future work on the topic would benefit from explicitly modeling the psychological processes involved in specific color tasks, in order to better understand how color terms affect the various components of cognition such as memory, decisions, and preferences.

Cross-References

- [Color Categorical Perception](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Psychological Color Space and Color Terms](#)
- [World Color Survey](#)

References

1. Pylyshyn, Z.: Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* **22**(3), 341–365 (1999)
2. Whorf, B.L.: *Language, Thought and Reality*. MIT Press, Cambridge, MA (1956)
3. Harnad, S. (ed.): *Categorical Perception: the Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
4. Bornstein, M.H., Korda, N.O.: Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychol. Res.* **46**(3), 207–222 (1984)
5. Kuhl, P.K.: Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**(11), 831–843 (2004)
6. The World Color Survey. <http://www1.icsi.berkeley.edu/wcs/>
7. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes, color perception is not categorical. *J. Vis.* **11**(12), 1–21 (2011)
8. Davidoff, J., Davies, I., Roberson, D.: Colour categories in a stone-age tribe. *Nature*. **402**(6724), 604–604 (1999)
9. Kay, P., Regier, T., Gilbert, A., Ivry, R.B.: Lateralized Whorf: language influences perceptual decision in the right visual field. In: Minett, J., Wang, W. (eds.) *Language, Evolution, and the Brain*, pp. 261–284. The

City University of Hong Kong Press, Hong Kong (2009)

10. Witzel, C., Gegenfurtner, K.R.: Is there a lateralized category effect for color? *J. Vis.* **11**(12), 1–25 (2011)
11. Kay, P., Kempton, W.M.: What is the Sapir–Whorf hypothesis? *Am. Anthropol.* **86**, 65–79 (1984)
12. Webster, M.A., Kay, P.: Color categories and color appearance. *Cognition.* **122**, 375–392 (2012)
13. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)
14. Witzel, C., Gegenfurtner, K.R.: Categorical facilitation with equally discriminable colors. *J. Vis.* **15**, 1–33 (2015)
15. Roberson, D., Hanley, J.R., Pak, H.: Thresholds for color discrimination in English and Korean speakers. *Cognition.* **112**, 482–487 (2009)

Efficient Coding

- [Environmental Influences on Color Vision](#)

Elastic Light Scattering

- [Rayleigh and Mie Scattering](#)

Elastic Scattering

- [Rayleigh and Mie Scattering](#)

Electrical Control Gear for Lamps

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Definition

Device(s) in the electric lamp circuit of gas discharge and solid-state lamps which limits the lamp current to the required value and ensures that the lamp can be started and once started can be operated stably over a more extended period.

The devices needed are called igniters, ballasts, and, for LEDs, drivers.

Igniters

Most of the gas discharge lamps have such a high internal resistance that they need a voltage peak, higher than the normal operating voltage, to initiate the discharge.

Igniters for Fluorescent Lamps

In most fluorescent lamp circuits, the electrodes are preheated for a few seconds before a high-voltage peak is applied across the lamp to initiate the discharge. Preheating of the electrodes facilitates the emission of electrons. Glow-switch starters are often used for this purpose. However, their harsh peak voltages have an adverse effect on lamp life. Electronic starters with exactly the same functions as the glow-switch starter, but with a better-controlled voltage peak, are used extensively. They always ignite after one ignition pulse, thus eliminating flickering during ignition. Because of the better control of the ignition peak, electrodes suffer less damage, so that lamp life increases.

Igniters for HID Lamps

Ignition of HID lamps is initiated by a high peak voltage without preheating of the lamp electrodes. Igniters for HID lamps are normally started using an electronic igniter that generates a series of high-voltage pulses of the required magnitude (varying for the different lamps between some 200 and 5000 V). The electronic circuit is so designed that these pulses cease after ignition has taken place. Because of differing requirements for ignition voltage, shape of voltage peak, and number of voltage pulses within a certain period, each type of HID lamp (and often also each different wattage) needs its own type of igniter.

Ballasts

If mains voltage is supplied to electrical devices, the current has to be limited; otherwise, it will

keep increasing until the device or circuit in which it is flowing breaks down. In many electrical devices, the internal electrical resistance of the device limits the current. These are devices with a so-called positive-resistance characteristic in which the current is automatically limited to a small range. Incandescent lamps have a positive-resistance characteristic and thus need no separate current-limiting device. Gas discharge lamps have a negative resistance characteristic, meaning that the unlimited current available will increase until the lamp breaks down.

Electromagnetic Ballasts

By introducing an external resistor in the lamp circuit, the current is stabilized. Simple resistors can thus be used as current-limiting ballasts for gas discharge lamps. They do, however, dissipate a lot of power. Instead, inductive coils are used, viz., copper wire wound around an iron core. They have the same effect as a resistor, but at much lower power losses in the ballast. The ballast also ensures that the lamp continues to operate despite the fact that twice during each frequency cycle of the mains voltage (in Europe 50 Hz and in the USA 60 Hz) the current is zero, the lamp is off and thus has to be reignited. An inductive ballast system helps to reduce the time the lamp is “off” during each zero passage of the current. This is because an inductive system shifts the phase between mains voltage and lamp current.

Until the 1980s, all ballasts consisted of inductance type of ballasts also called “electromagnetic ballasts” because of the electric and magnetic fields their coils generate. Since then, more efficient electronic ballasts have been developed.

Electronic Ballasts for Fluorescent Lamps

Electronic ballasts for fluorescent lamps electronically transform the sinusoidal 50 Hz mains frequency into a square-wave voltage of higher frequency, hence their name HF electronic ballasts. The high-frequency voltage lies between 25,000 and 105,000 Hz (viz., 25–105 kHz). HF electronic ballasts work on the same inductive principle as conventional electromagnetic ballasts. At the higher frequencies, much smaller coils with correspondingly smaller losses can be

used than in conventional electromagnetic ballasts. When a fluorescent lamp operates on a frequency higher than some 10 kHz, the efficiency with which the radiation is created increases. The more efficient light production combined with the lower ballast losses means that the system efficacy of fluorescent lamps operated at high-frequency increases by some 20–25% compared to that of the same lamps operated on electromagnetic ballasts.

The electronic ignition function is incorporated in the ballast. With the electronic operation, the exactly required starting voltage pulse can be accurately supplied to the lamp. This minimizes electrode damage and so considerably increases lamp life. Relative to lamps operated on electromagnetic ballasts, lamp life is increased by 30–50%.

Electronic Ballasts for HID Lamps

Electronic ballasts for HID lamps have become available for some types of HID lamps. In contrast with HF electronically operated fluorescent lamps, there is hardly anything to be gained in terms of improved efficiency in the light-creation process by operating HID lamps on an electronic ballast. Nevertheless, energy savings are obtained, because the power losses in electronic HID ballasts are lower. The advantages of electronically operated HID lamps often lie in the better and easier control of certain lamp properties.

Drivers for LEDs

Instant starting is no problem with solid-state lamps, but they are low-voltage rectifiers that allow current to pass in one direction only. This means that the AC mains supply has to be transformed to low voltage and then rectified into a DC supply. Although a solid-state light source has a positive resistance characteristic, the voltage-current dependency is exponential in the area of operation. Small fluctuations in supply voltage, therefore, cause large variations in current that can damage the light source. A simple series resistor in the electrical circuit stabilizes the current to create, in fact, a “constant” current supply.

In practice, only miniature indicator LEDs use such a resistor. Because so much energy is lost in resistor-type drivers, all high-power, high-brightness LEDs employ electronic drivers that provide for the transformation from high voltage to low voltage, for rectification, and for the constant current supply.

Power losses in electronic drivers vary, according to their quality, between something like 10% and 30%, with really poor-quality drivers, as much as 50% of the nominal wattage of the LED itself.

Cross-References

- ▶ [High- and Low-Pressure Sodium Lamp](#)
- ▶ [High Pressure Mercury Lamp](#)
- ▶ [Metal Halide Lamp](#)
- ▶ [Tubular and Compact Fluorescent Lamp](#)

Electrodeless Low-Pressure Mercury Lamps

- ▶ [Induction Lamp](#)

Elementary Colors

- ▶ [Primary Colors](#)

Emergency Lighting

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
Netherlands

Definition

Lighting to ensure the safety of users and visitors of a building when the normal lighting fails.

Security Lighting Standards

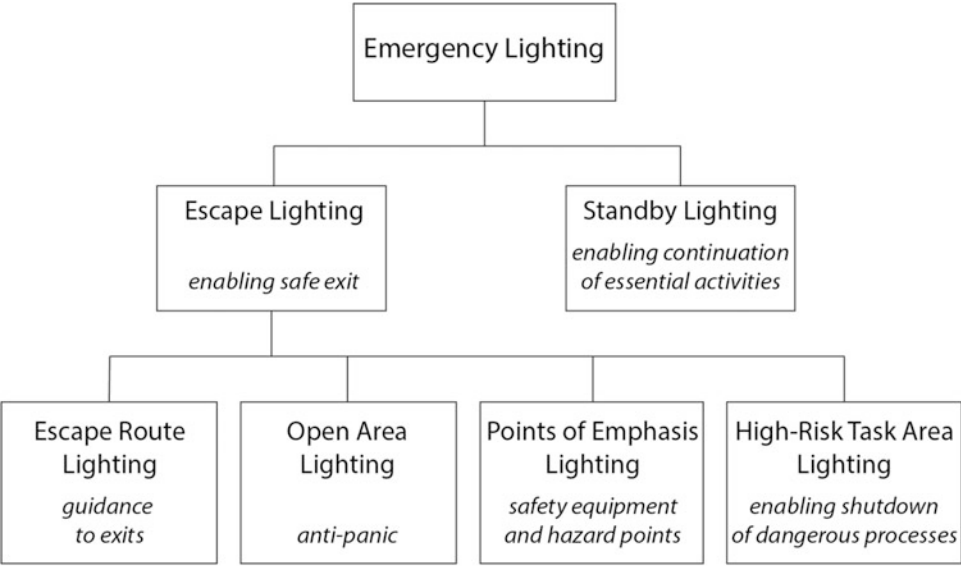
The most important objective of emergency lighting is to ensure the safety of users and visitors of a building when in the case of a calamity the normal lighting fails. Emergency lighting, therefore, uses a power source that is entirely independent of the source supplying the normal lighting. In most countries, the emergency lighting of buildings is regulated in laws. These laws may be aimed at those responsible for the building and building installations or those responsible for the people in the building. In many countries, these laws refer to national or international standards for emergency lighting. Such standards are usually split in, on the one hand, standards for specifying the emergency lighting quality and, on the other hand, for specifying the quality of the products used in the emergency lighting installation. The emphasis here is on the emergency lighting quality specifications.

The International Lighting Commission, CIE, produced in 2007 an ISO/CIE standard that specifies lighting requirements for emergency lighting [1]. Many national and regional standards are based on this standard. The 2013 emergency lighting standard of the European Standardisation Commission, CEN, is an example of such regional standard [2]. This standard has incorporated some clarifications and improvements where inconsistencies existed in the 2007 ISO/CIE standard. Therefore, here the more recent European standard is used as a basis for recommended lighting values.

Emergency lighting levels specified in the next sections are maintained lighting levels and do not take into account indirect light contributions so that the minimum effect of the emergency lighting is not affected by changes in the reflectances of room surfaces, for example, by repainting or redecoration.

Categories of Emergency Lighting

Emergency lighting comprises lighting that facilitates the evacuation of a building in an emergency, as well as standby lighting enabling the continuation of activities of vital importance



Emergency Lighting, Fig. 1 Different categories of emergency lighting with, in italic characters, the objective of lighting

during a power failure (Fig. 1). To facilitate evacuation, persons should be able to orientate themselves in their environment, to locate an escape route which is indicated with discernible signs and safely follow that escape route to the exit. As shown in Fig. 1, escape lighting can be subdivided into escape-route lighting, enabling a safe exit, anti-panic open-area lighting, lighting of points of emphasis such as places with safety equipment, and high-risk task area lighting enabling the shutdown of dangerous processes. Safety signs used for guidance in an emergency should have internal or external lighting that makes them conspicuous and legible from appropriate distances.

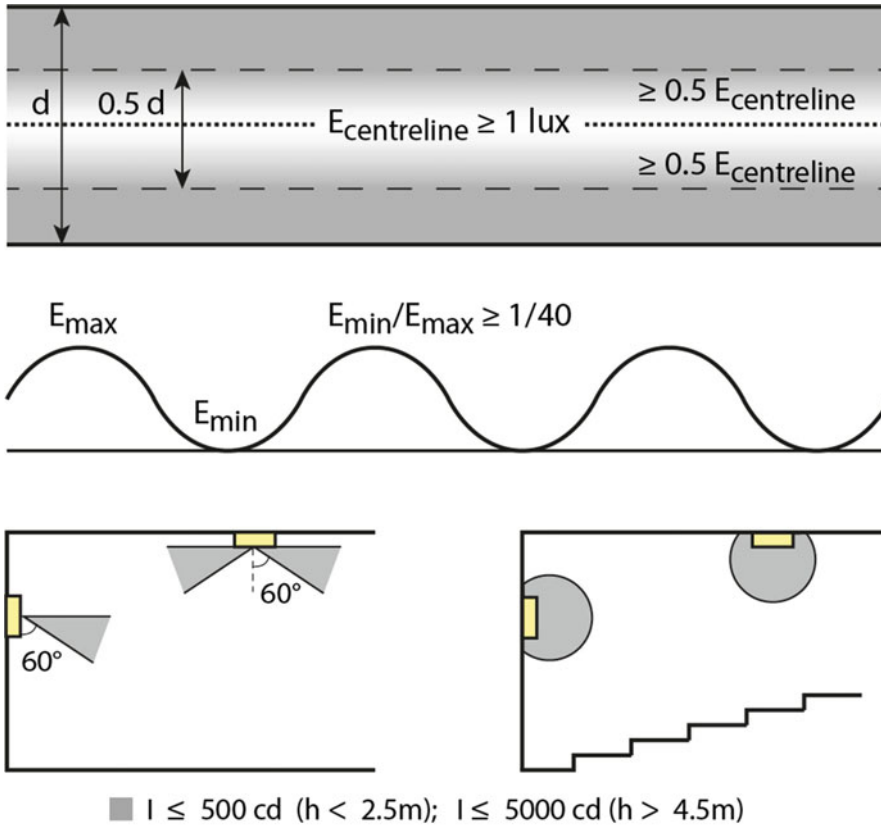
Escape-Route Lighting

The emergency plan designated escape routes, marked with safety signs using pictograms and colors, should have lighting that identifies the escape route and enables the safe use of it. For this purpose, the center line of the escape route should have illuminance values at floor level, along each point of the line, of at least 1 lux. For escape routes up to 2 m in width, a band centered around that central line, consisting of at least half

the width of the escape route, must have at all points at least an illuminance of 50% of the illuminance value provided on the central line (Fig. 2, top). Broader escape routes can be treated as a number of 2 m wide strips or as anti-panic open areas.

If the diversity of the lighting, i.e., the difference between the minimum and maximum illuminance values, is too high, the higher illuminance values may affect visibility in the areas with lower values. To restrict this adverse effect, the diversity of the lighting (E_{min}/E_{max}) must be better than 1/40 (Fig. 2, middle).

Disability glare should be limited. The ISO/CIE and CEN standards specify maximum luminous intensities for the zones shown in gray in Fig. 2, bottom. For equal-level floors, the luminous intensities must be limited within the zone 60° to 90° from the downward vertical, whereas for non-level escape routes, such as stairs, the limiting values apply to all angles (see Fig. 2, bottom, again). The luminous intensity limit of 500 cd as shown in Fig. 2 is valid for luminaire mounting heights smaller than 2.5 m. The standards specify larger values for a series of larger mounting heights. The largest intensity tolerated



Emergency Lighting, Fig. 2 Escape-route lighting requirements for routes up to 2 m in width. Top, illuminance levels; middle, illuminance diversity; bottom, zones

in which luminous intensities have to be limited to restrict disability glare. Values based on [2]

in the standards is 5000 cd for mounting heights larger than 4.5 m.

The emergency escape-route lighting must reach 50% of the required illuminance within 5 s and 100% within 60 s. The escape-route lighting should continue to provide lighting for at least 1 h.

Escape lighting should not only make the route itself visible but also the safety signs should be discernible. For this reason, the color-rendering index R_a of the light sources used for the emergency lighting should be at least 40.

Open-area lighting: Open-area lighting is also referred to as anti-panic lighting. The objective, apart from helping to avoid panic, is to facilitate reaching a place from where an escape route can be easily found. The horizontal illuminance at floor level must at least be 0.5 lux, and the

diversity ($E_{\min}/E_{\max}$) should be better than 1/40. Disability glare should be limited just as specified for the escape-route lighting. Also the start-up time, duration, and color-rendering index of the light sources should be the same as for the escape-route lighting.

Points-of-emphasis lighting: Points of emphasis are points where an escape-lighting luminaire has to be placed (less than 2 m from the point). If relevant, such a luminaire can be part of the escape route or anti-panic open-area lighting. The most critical points of emphasis are:

- Exit doors intended for use in an emergency.
- Each flight of stairs or any other change in level.
- Each change of direction and intersection of corridors.

- First-aid posts, call points, and points with escape equipment for the disabled; first-aid boxes and firefighting equipment must have emergency lighting providing a vertical illuminance of at least 5 lux.
- Disabled refugee points.
- Safety signs that are part of the emergency plan.
- Outside the final exit at a place of safety.

High-Risk Task Area Lighting

The objective of this category of emergency lighting is enabling the shutdown of dangerous processes; these may range from short-lasting hazards like operating an electric saw or long-lasting hazards like a metal or glass oven with the danger of hot liquid pouring out. General emergency lighting standards may not cover all potential risks. In the analysis phase of the emergency lighting design process, it should be carefully checked with the responsible person for the processes whether the standard is relevant enough.

Typically, high-risk task areas should at least have an illuminance on the task area of 10% of that required for the task under normal conditions, but not less than 15 lux. The illuminance uniformity, E_{min}/E_{av} , should be better than 0.1.

The lighting must reach 100% of the required illuminance within 5 s unless the process is identified as one requiring a shorter start-up or permitting a longer one. The lighting should continue to provide lighting as long as the risk lasts.

Disability glare should be limited just as specified for the escape-route lighting, and the color-rendering index should also be the same, i.e., at least 40.

Standby Lighting

The objective of standby lighting, also called substitute lighting, is enabling an uninterrupted continuation of activities of vital importance during a power failure. The vital importance can have a safety but also a technical or economical reason. An example where standby lighting is required is the hospital operating theater. The lighting requirements of the standby lighting have to be

discussed with the responsible persons for the activity. The lighting requirements are often the same as the requirements under normal operating conditions.

Emergency Lighting Installation

The emergency lighting installation has its own power supply system, completely separate from the mains power supply of the normal lighting. The system can be centralized or decentralized. In a centralized system, one central power supply unit powers the emergency lighting luminaires. In a decentralized system, the individual emergency lighting luminaires are equipped with a chargeable battery making it self-contained luminaires.

Centralized System

A centralized system uses a set of accumulators or batteries of sufficient power to start and maintain the emergency lighting through an own cabling network. The units are located at a central location appropriately protected against disturbing external influences. A non-break generator can be used to take over from the batteries to extend the period that the emergency lighting stays on. The cabling network has to fulfil stringent requirements, in particular as far as fire protection is concerned. Centralized systems are so-called nonpermanent systems in the sense that they are normally switched on only in an emergency.

The emergency luminaires can be given a double function by powering them from the mains under normal conditions and integrating them in the normal lighting. In an emergency, they continue to receive power, now from the emergency power system. Of course, such integrated luminaires must be of the same type (visually and light-technically) as the other luminaires used under normal conditions. For escape-route lighting, this may be not realistic, but for the open-area lighting, it may be a possibility.

Regular tests to check the proper operation of the emergency system can easily be carried out with central systems.

Decentralized System

In this system, the individual emergency lighting luminaires have an individual power supply in the form of a chargeable battery. The luminaires thus function self-contained or autonomous. The luminaire itself incorporates the functions of power supply, battery charging, mains failure detection, and illumination. The battery is charged under normal conditions by the standard mains power. Self-contained luminaires can be used individually or in groups. Luminaires can be executed as permanent or nonpermanent luminaires. Permanent luminaires provide lighting also under normal conditions. This can, for example, be applicable for the escape-route signalization. In this way, the users of the space can get acquainted, under normal conditions, with the emergency escape routes. At night, they can serve, partly or entirely, as security lighting.

An important advantage of decentralized systems using self-contained luminaires is that the system cannot be affected by faults in the central power supply or cabling network. In the case of a calamity, especially the latter can become a reality. If there is a power defect in only one or a limited number of groups, the self-sustained luminaires in that group or groups respond. Extensions of the emergency system are easily made with a decentralized system. A disadvantage of decentralized systems is that the regular tests of the proper functioning of the system are time-consuming for the maintenance staff and may interrupt normal activities at the location of the luminaire. Self-test emergency luminaires can replace the tests carried out by the maintenance staff. When a defect is detected in the self-test carried out by the luminaire itself, it displays a visual warning.

References

1. ISO/CIE (2007) Publication ISO 30061:2007(E)/CIE S 020/E:2007 Joint ISO/CIE Standard: emergency lighting
2. CEN: EN Standard 1838:2013 Lighting Applications: Emergency Lighting. European Committee for Standardization, CEN, Brussels (2013)

Emotional Color Effects

- [Psychological Color Effects](#)

Environmental Chemistry of Synthetic Dyes

- [Colorant, Environmental Aspects](#)
- [Coloration, Hair](#)

Environmental Color Design

Galen Minah and Antal Nemcsics

Department of Architecture, College of Built Environments, University of Washington, Seattle, WA, USA

Synonyms

[Color dynamics](#); [Contextual color design](#); [Dynamic color design](#)

Definition

Environmental color design involves the use of color in order to configure it as a more beautiful, usable, and informative component of the environment that allows theoretical and practical activities. Color design is a relevant aspect of environmental design, which includes such disciplines and practices as landscape design, urban planning, architecture, interior design, industrial design, engineering, graphic design, textile and fashion design, etc., all of them aimed at a transformation of the natural environment with the purpose of adapting it to human life.

Antal Nemcsics: Deceased.

The Function of Color in Space Experience

Man living in a built environment, the user of a built space, is protected from the tribulations of nature; he utilizes the services of his surroundings and enjoys comfort from these services. In addition to its actual measurable properties required for the physical and biological existence of man, built environment has other qualities too.

Built space acts upon man in several ways: by the proportions, the relationship, and shape of its elements, the order of forms, by surface appearance and colors of the elements, by the relation between space proportions, by the expression of function, by the relation between the expressivity of function and the function proper, and by the shape and color associations expressing function. This effect materializes as an emotional experience of the actual space, and space sensation. Space sensation is an experience about a given space, an accomplishment of one's own personality.

The Function of Color in Space Perception

Space perception is a complex process to which several sensory organs contribute. Among them, visual and auditory stimuli and those arising from motion in space are the most important. All these add up to a space stimulus eliciting the perception of space.

Space stimulus is elicited by measurable and tangible real space, composed of space elements as well as of correlations between shapes and surface appearances, all describable by physical magnitudes. Most of the information about the objective correlations, shapes, and surface appearances of space elements is obtained by reflection, absorption, or transmission of light by the surface of the element that delivers visual stimuli to observers.

Assuming that the surface appearance of space elements is of the same finish, texture, and color, and that the elements are illuminated from the same direction, with the same intensity and

spectral distribution, then due to visual and motional parallaxes, overlapping, line and air perspective, and light-shadow effects, a space perception with a linearity directly proportional to that of the change of the real space is elicited.

Color identity as a condition means that light incident from the surfaces into the eyes has the same wavelength; that is, color sensation is the same throughout, and also, that for the same angle of incidence, the ratio of the quantity of light incident on and reflected by surfaces is the same everywhere, and in the reflected light incident on the eyes, the ratio of complementary radiations, hence saturation, is felt to be the same throughout.

To examine the function of color, let us assume that stimuli arriving into the eyes come from the elements of such an objective space where dimensions, proportions, and relations of the elements do not permit overlapping and the interpretation of line perspective relations; further the onlooker does not move in space, missing the help from laws of motion parallax in space perception. If these conditions are met, and in addition direction, intensity, and spectral energy distribution of light within the space are constant and the former restriction of equal hues, saturations, and lightnesses of surface colors holds, the objective space can be judged only by evaluating the perceived color sensation differences.

Intensity differences of the stimuli emitted by space element surfaces and reaching the eyes permit, first of all, to decide on the spatial position of the light source, then, from hue, saturation, and lightness differences of space element surfaces, on the distance of space elements from the onlooker, hence on the space itself.

It is known by experience that the more remote an object, the more hue component of the color sensation generated by its surface is shifted to hues of shorter wavelengths, its saturation toward achromatic colors, and that its lightness component varies as a function of the two other components and of the position of the light source. This experience helps the space sensation although its significance can really be perceived only if the former condition of color identity is abandoned. In reality, this is always the case. With space elements painted different colors, it cannot be

decided anymore which element is the closer and which is the farther away. Orange and red, even if in reality more distant, are felt to be nearer than blue or green. Saturated colors are felt to be nearer than are unsaturated ones. Very dark surfaces emit very little or no stimuli to the eye, so that these are not sensed, rendering space perception impossible.

The Role of Color in Expressing the Function of Space

Color contributes to space sensation also by expressing the function of space. Function of the built environment is a demand raised to social level. Structural relations in a system composed of man and the elements of his environment are defined by a complex function having three components: utility function, aesthetic function, and informative function. This section explains how color – color stimulus and color perception – contributes to the realization or expression of these functions.

Environment is the scene of human activities, serving human demands. Much of these demands refer to the utility function of environment. Built environment is required to protect from the rigors of weather, to endure dynamic forces generated by human machinery, to protect from such factors as excessive temperature fluctuations, intense noise, and other factors from working processes detrimental to health. A recent requirement is feeling of comfort in one's milieu so as to stimulate the development of mental and physical abilities.

Color has a significant function in meeting these demands. Due to its psychophysical and psychosomatic effects, it may raise blood pressure or change the composition of blood and gastric juices. Color can make one feel healthy or ill. A person in an environment of preferred colors feels better; his/her ambition to work increases. Some colors favor concentration; others cause deconcentration [1].

Just as anything else, built space and all its elements are separable unities of content and form. Environment fulfills its aesthetic function if it expresses its utility function in conformity

with the unity of content and form, where the utility function is the content, while form is expressed by shape and color of environmental elements. Since the content of objects in the environment, let alone in the built environment, is its function, the built space and objects within its content can only be grasped and fully expressed by means of their proper functioning and operation. Practical and spiritual components of the function are interdependent. Even the remark may be risked that aesthetic design of an object or built space is impossible when ignoring its functions. As a conclusion, there are no aesthetic prescriptions of general validity.

In designing color relations for a built environment as a human creation, it is also a question of what importance is attributed to practical functions of the environment for human life in general. Every work and activity is linked to emotions, thoughts, and ideas, therefore every object, tool, or built space demands its share of these mental, emotional, ideological threads, in conformity with its role, significance, and function in one's life. Colors of the built space as elements of form in the couple content and form are made necessary by the sensation of function, giving rise to a harmonious sensation of the indissoluble unity between content and form in human consciousness. Of course, the sight of some color complex may cause aesthetic pleasure, but detached from the content of space, i.e., from its function, this pleasure lacks the effect of complete space experience.

Those who wish to express the message of built space have to know about relations between environmental structures, i.e., about the so-called compositional relationships, in order to be able to create proper relations between forms and color perceptions. These relations comprise color harmonies. Thus, the design of space sensation also exploits color harmonies in this space.

Informative functions of space are features that interpret the functions of the environment and its elements and explain how to use and operate these elements. A significant part of the informative functions of the environment are borne by chromatic information. According to their message, chromatic information may be interpreted either as logic or as aesthetic information. Both kinds of

information are borne by the same elements, but every form of message has its own structure. Their characteristics are determined partly by differences in the visual system, complexity, and structure and partly as psychic differences between their communication content. Information content is transmitted by highlighting, contracting, and grouping some visual symbol elements in the informative surface or space, while omitting others. A color group draws attention when it is clear cut and its structure is easily intelligible.

Chromatic information of a logic nature, i.e., the various standardized color codes, are practical tools that appeal to a logical mind. They transmit messages and serve also to influence observers in their decisions and control their attitudes and behavior.

Aesthetic chromatic information is primarily emotional expression of inner conditions and is expected to have mental and emotional effects by commonly accepted semantics. By their operative and recording functions, visual codes are not only bearers of the meaning of the content of built space and its social concept but also expressions of the approach and culture typical of the creative subject. Chromatic in-built space information necessarily and conveniently takes the form of color harmony relations. This is why it is of prime importance to examine color harmony relations.

The Color Dynamics of Color Theory and Practice of Environmental Planning

Nowadays color dynamics is regarded as a dual activity. One side is the disclosure of man-to-color of complex man-to-colored environment relations and the elaboration of methodologies for the design of colored environment. The other side is the utilization of these findings in environment design practice. These activities involve the collection and systematization of knowledge on relations of man, color, and built space offered by different disciplines as well as to devise and realize research to fill the gaps. This activity has taken momentum worldwide, whereby the science of color dynamics came into being.

Color dynamics as a new science is concerned with the relations between the surface appearance of environment and environmental elements and man living in this environment. It studies the interrelations of color, man, and environment. Thus, color dynamics as a science is a complex of theoretical and practical activities directed toward the disclosure of objective relations between man and colored environment, as well as toward a conscious environment color design.

Rather than a collection of everyday experience, color dynamics is a science. Although it investigates and processes the spontaneous, intuitive transformation of the environment by individuals, it handles its information by scientific methods, applying scientific methods in environment color design. It has been proven both theoretically and practically that it is possible to conduct these activities by exact, scientific methods.

Environment design – including any architectural activity – has increasing access to results of this new science. Its practical application helps the built environment to cope better with its function, to be more beautiful and more sophisticated. It helps humans to expand mental and bodily abilities, to compensate for harmful effects of the overwhelming industrialized environment, to develop an adequate space perception, and to understand spatial relations and correlations between spatial processes. Conscious application of colors is expected to direct and orient man between multiplying and often depressive environmental hazards.

The science of color dynamics has five different but inseparable branches; the achievements of each are interdependent. Knowledge and relations amassed by other sciences are collected and purposefully systematized, and its special research problems are based on this foundation.

The fundamental problem of color dynamics is to find relations between color sensations, to develop an aesthetically uniform color space and a color system fairly approximating to it, and to introduce a new system of color coding suitable for practical color design.

The second group of problems is concerned with man to color relation independent of the

environment. This involves color composition problems in connection with the processes of color vision. Such problems are, for instance, stimulus thresholds and difference thresholds, color adaptation, color constancy, color contrast, color preference, color association, and the psychosomatic effects of color.

The third group of problems includes the complex relation between color, man, and built space, including problems of color and space, color and mass, color and form, color and texture, color and function, color and illumination, offsetting harmful environmental effects by color, and finally the social functions of color.

The fourth group of problems is related to color harmony research, the establishment of color composition relationships for use in practical color design: the determination of levels and parts of the concept of color harmony, of the fundamental and accessory conditions of eliciting color harmony sensations.

The last, fifth, group of problems is the development of the most effective methods of color design, the best way of incorporating the finding of color dynamics into practice. Statements are made exploiting practical observations obtained from realized color designs.

Color environment design activities, as all other design activities, constitute a continuous battle between prevailing conditions, demands of different trends, and artistic intentions. This battle may express itself on many different ways depending on the knowledge base and personality of the designer. Color dynamics has no intent to provide design recipes; rather it shows a method to provide possibility for considering different conditions and enforcing different demands [2].

Color Expression in the Built Environment

Finding a basis for using color as a concept in design is possible through the perceptual experience of color and form. Figure/ground recognition in a visual field is based upon perception. Figure/ground also implies a hierarchical relationship.

Color and form are codependent in built form. This form can be holistic as it relates to an urban environment, or it can be an individual building. Color plays an expressive role in both. In the city, color has the power to create unity in built form by the use of a dominant color in the building materials. This visual unity can be a characteristic of a part, or district, within an urban environment, or it can be a characteristic of the whole. Unity defines a sense of place and gives identity. It also characterizes an urban fabric or background that can define the public spaces and serve as a backdrop for significant figural structures. The architecture of this background is critical to the designer's goal of creating spatial definition and unity in the plan of a city.

Another role for color expression in the built environment is the color choice for individual buildings that become prominent visually in a physical setting. These are defined as figural colors, and they will achieve this visibility through figure/ground contrasts. The atmospheric conditions and the distance from the viewer will also be a determinant in how effective the color choices are in achieving high visibility.

A third tool for color expression in the physical environment is a strategy for defining space not with walls or edges but with a repetitive use of figural color throughout the space. This strategy uses color constellations formed by elements of similar color in the environment that create a spatial matrix when seen as a whole. Constellations can define urban space in two ways: by experiencing these elements of color as a constellation when viewed as a group and through memory when one views a single-color element from a pedestrian view and constructs the whole conceptually.

Background Architecture

In urban contexts, background is that median color/form of the built environment which most conforms to the overall urban fabric and contributes to its visual unity. Background is also the structure or mass that forms and defines the spaces of importance within the city. The color of this background fabric is a component in the overall visual field. The color of building surfaces comprises a large portion of this field, and these colors



Environmental Color Design, Fig. 1 Dark towers as figural color in the central business district, Seattle, Washington

can be a powerful means for defining unity and consistency. A variety of figural form and color can exist as background in a city through repetition of similar elements distributed throughout the urban fabric. In some cities, such as San Francisco, the background is relatively uniform in hue and saturation. Guanajuato, Mexico, however, has many buildings with highly saturated figural colors throughout the city and maintains its visual unity despite these foreground colors. There are two important roles for color/form in this background architecture. The first is the provision of a context for contrast to structures that are figural and highly visible in the city. The second is one of spatial definition. Background architecture becomes the edges and walls that define public spaces such as streets, avenues, civic plazas, and parks, and their organization represents the hierarchies and communal goals inherent in the plan of a city.

Figural Color

Three perceptual experiences of color have direct relationship to questions of architectural color and

form. These are color contrasts that create figure/ground juxtapositions, the spatial effect of color, and the atmospheric color effect. Of these, figure/ground is important conceptually because it represents a hierarchical structure to which expressive values can be assigned. Figural structures in an urban environment are defined by their mass and height, their form, their position or siting within the city, and their color. In an urban context there are colors that make some buildings figural in contrast to their background. Background colors in many modern cities tend toward low blackness ($s = 10\text{--}30$) and low chromaticness ($c = 10\text{--}40$). Building colors that approach maximum whiteness, blackness, or chromaticness will usually be seen as figural in an urban environment [3]. The climate and atmosphere will affect how figural these colors appear. In most weather conditions, white and saturated colors are prominent. Dark colors will tend to be less figural in overcast or humid conditions and appear lighter [4]. Figural color in cities creates visibility for buildings, and as a tectonic tool this visibility can denote status and provide meaning to a civic community (Figs. 1, 2, and 3).



Environmental Color Design, Fig. 2 Saturated *yellow* as figural color on Queen Anne Hill, Seattle, Washington

Color Constellations

In an urban context, there are examples where figural colors and background colors become joined as a means for defining space within the city. In a two-dimensional color field where many colors are visible, color constellations are groups of similar colors recognized as a pattern or cluster. In a three-dimensional color field such as an urban environment, colors that are similar in hue and saturation can also be perceived as a cluster of repetitive color elements and form color constellations [5]. The area within this array of similar colors can be designated as an urban space, defined not by edges or walls but by a repetitive matrix of similar colors. An example would be the Parc la Villette in Paris [6], where large red architectural sculptures are placed on a grid throughout the precinct of the park. From an aerial perspective, the spatial matrix formed by these red structures is apparent. From a pedestrian viewpoint, only a few of the structures are visible at any one time, although a memory of the precinct as a

whole initiated by the individual red sculptures is always present.

Many city maps are organized by color sections to show important landmarks and institutions, shopping/retail districts, public open spaces, nightlife areas, etc. Many of these nodes and districts are repetitive and dispersed throughout the city. The map represents patterns of color clusters, like a constellation, and a large amount of information about the organization of the parts of the city is made visually accessible and comprehensible. Color in the physical environment has the potential for providing the same visual comprehension. Many cities have zoning in these districts that specify heights of buildings, density, signage and lighting, etc. that help in achieving the goal of comprehensibility from a pedestrian's view of the city, but color is rarely a consideration. One example, however, is in Washington, D.C., where the important governmental and institutional buildings are white with the background architecture in darker hues. These important buildings and

Environmental Color Design, Fig. 3 *White* as figural color, denoting status, Schiers, Switzerland



Environmental Color Design, Fig. 4 Color constellation at Parc La Villette, Paris, France

landmarks form a constellation of white figural structures located in the vicinity of the Mall in the center of the city. As in Parc La Villette, the view of

any of the white buildings recreates an image of the group of figural buildings around the Mall, and any white building near the area becomes a conceptual

part of this collection of significant structures (Fig. 4).

Cross-References

- [Appearance](#)
- [Color Constancy](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Color Preference](#)
- [Functionality of Color](#)

References

1. Frieling, H.: *Farbe in Kultur und Leben*. Battenberg, Munich (1963)
2. Nemcsics, A.: *Colour dynamics*. Environmental colour design. Ellis Horwood, New York (1993)
3. Minah, G.: Blackness, whiteness, chromaticness: formulas for high visibility in the modern city. In: Hansuebsai, A. (ed.) *AIC 2003 Bangkok, Proceedings*, pp. 26–30. The Color Group of Thailand, Bangkok (2003)
4. Minah, G.: Figural color in the Seattle cityscape. In: *AIC color 97, Proceedings of the 8th Congress of the International Colour Association*, pp. 883–887. The Color Association of Japan, Tokyo (1997)
5. Minah, G.: Color constellations in the Seattle cityscape. In: Chung, R., Rodrigues, A. (eds.) *AIC 2001, the 9th Congress of the International Colour Association*, pp. 146–149. SPIE, Bellingham (2001)
6. Tschumi, B.: *Cinégramme folie: le Parc de la Villette, Paris nineteenth arrondissement*. Princeton Architectural Press, Princeton (1987)

Environmental Influences on Color Vision

Kassandra R. Lee and Michael A. Webster
Department of Psychology, University of Nevada,
Reno, NV, USA

Synonyms

[Color adaptation](#); [Color and culture](#); [Color constancy](#); [Color signals](#); [Efficient coding](#); [Evolution](#); [Individual differences](#)

Definition

It is evident that color vision evolved to inform organisms about their environment and thus must be shaped by the environment the organism is in. The color properties of the world are a potent factor affecting not only the basic mechanisms of color vision within a species but also how they are calibrated and fine-tuned within the individual. As a result, many aspects of the physiology of color coding and the phenomenology of color experience can be predicted by looking outside the observer – to analyze the signals available in the color environment. The following provides a number of specific examples of how knowledge of the color characteristics of the environment has helped to account for some of the known properties of our color vision.

Environmental Signals and the Evolution of Color Vision

A central principle in understanding the design of sensory systems has been coding efficiency – or how to transmit the most information about the stimulus using the least resources. The assumption that vision maximizes coding efficiency provides a powerful link between the properties of the stimulus and the properties of the observer [1] and has led to a surprising number of insights about the early stages of color vision and how they are optimized for the color characteristics of the environment. For example, natural color spectra (of both illuminants and surfaces) typically vary gradually with wavelength and can be represented and discriminated based on a small number of spectral samples [2]. This may in part be why color vision is typically low dimensional (e.g., based on three classes of photoreceptors in trichromatic observers). The fact that spectra can be represented by a small number of dimensions has also been a central insight into understanding the capacity and limits of color constancy (e.g., the ability to perceive that the color of a surface has not changed even when the light falling on it has).

Scenes are composed of collections of many spectra, and these color distributions have

characteristic properties. For example, most luminances and chromaticities tend to be near the average, or gray. This property has been used to predict how neurons respond to different intensity or color levels in the image. To code these efficiently, responses should change most rapidly near gray, where signals are more frequent and thus require better discrimination, and asymptote at the extremes (e.g., highly saturated colors), where signals are rare. Thus the color statistics of images can directly account for the sigmoidal response functions of visual neurons and for the fact that we are best at discriminating small color differences near gray [3].

Color vision requires comparing the differences between the responses of receptors with different spectral sensitivities. Because the sensitivities are broad and overlapping, and because as noted spectra vary gradually, the signals within the different cone classes are highly correlated. This redundancy can be removed by recoding the cone signals into “luminance” mechanisms (e.g., in cells that receive excitatory inputs from different cones and thus sum their inputs to represent luminance) and “chromatic” mechanisms (e.g., in cells that instead receive excitatory inputs from one cone type and inhibition from another, so that the response depends on the difference or ratio of the cones’ signals). Thus redundancy reduction provides a rationale for how the cone signals are reorganized in cells in the retina. Finally, the overlap in the cone sensitivities also means that the differences between the cones (which convey color) are many times smaller than the absolute responses (which convey luminance). If our vision depended on these raw cone signals, then the world would vary much more in brightness than in color. However, a further prediction from coding efficiency is that each channel uses its fully dynamic range to code the range of available signals [3]. As a consequence we are much more sensitive to chromatic differences than luminance differences (when equated for cone contrasts), and the range of brightness and color in the world appears perceptually balanced.

The preceding examples illustrate how studies of the color environment have provided insights into general properties of the mechanisms of color

vision. For many species the environment also holds special color signals that are important to detect, because they signal food or mates, and these signals have given rise to more specialized adaptations. For example, most mammals are dichromats, yet humans and other old-world primates evolved a third class of photoreceptor through an evolutionarily recent duplication of the gene coding their longer-wave photopigment. The extra dimension of color afforded by their trichromacy has been shown to be well tuned for detecting ripe fruit among foliage and also for judging the subtle variations in skin tones that signal the health and emotional state of conspecifics [4].

Color Appearance and the Color Environment

Conventional models of color appearance hold that the signals from the three classes of cones are recombined to form three perceptual channels that signal red vs. green, blue vs. yellow, and bright vs. dark. A central issue in color science is how these perceptual dimensions are encoded by neurons of the visual pathway and why particular hue directions are special. Here again the color characteristics of the environment are thought to play a central role.

According to the standard color-opponent model, any stimulus is represented by how it differs from gray (e.g., lighter vs. darker or redder vs. greener). But what establishes gray? The spectral sensitivity of the observer varies in capricious ways depending on factors such as the level of screening pigments or the relative numbers of the different types of cone receptors and thus must be adjusted to match “gray” in the environment. This gray likely corresponds to the average spectral stimulus the observer has been exposed to. Normalizing color responses for this average probably happens at many levels of the visual pathway and at many timescales. For example, there is increasing evidence for distinct processes of short-term and long-term adaptation to color at multiple stages of visual processing [5]. Color adaptation initially arises in the photoreceptors,

which rapidly adjust to changes in the average color or brightness of the scene (e.g., becoming more or less sensitive when the light levels get lower or higher, respectively). These sensitivity changes rescale their relative responses so that the current average color tends toward gray, an adjustment known as von Kries adaptation, and this is thought to be an important initial component in color constancy (since it tends to remove or discount changes in the average color of the scene introduced by changes in the illuminant). Similarly, opponent channels must be calibrated so that the opposing cone inputs are weighted so that their null points (e.g., where the signal is neither red nor green) correspond to a neutral gray. Sensory adaptation thus plays a critical role in calibrating color vision so that it is correctly adjusted to match the observer's environment. (Importantly, these adaptive responses are not unique to color, but probably influence all aspects of perception [6].)

Supporting this, many lines of evidence show that the stimulus that appears gray depends very little on the observer's actual sensitivity to wavelength. For example, individuals vary widely in the density of the lens pigment, which also steadily builds up with age. This pigment selectively absorbs light at shorter wavelengths, and thus the light reaching the retina of an older observer will have a different spectrum from a typical young observer. Despite this, the physical stimulus that appears white remains constant across the lifespan, because color mechanisms are continuously recalibrated so that they are adjusted to the average spectrum [7]. Receptors in the fovea at the center of vision are also selectively screened by a second inert pigment known as macular pigment, which further filters light at short wavelengths. Thus the fovea and surrounding periphery receive very different stimulation on average. Yet the physical stimulus that looks white again remains largely constant across the visual field, suggesting that each part of the eye can locally adapt to set the achromatic point. Patients undergoing cataract surgery provide a natural experiment for observing these adaptation processes. Cataracts represent an extreme form of lens brunescence, blocking much of the short-

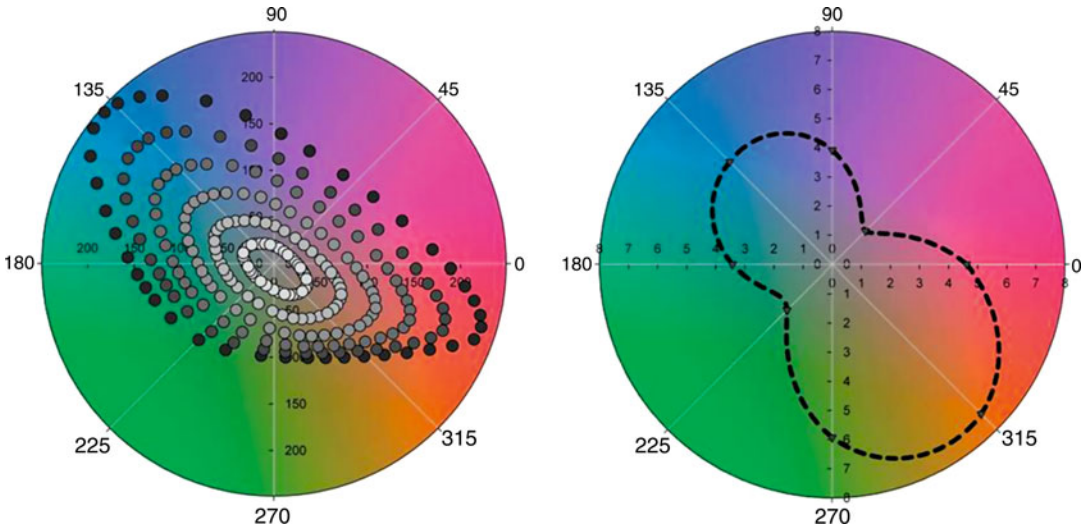
wavelength light. Yet the stimulus that appears white is within the normal range for many patients. When the cataract is replaced with a clear lens, the sudden increase in short wavelengths causes the world to appear too blue. However, achromatic settings almost immediately begin to renormalize and continue to slowly drift back to their prior levels over a period that may last for months [8].

Analogous arguments have been advanced to account for the special nature of red-green and blue-yellow as the primary dimensions of color appearance. As noted above, red may be special because it signals ripe fruit or blood. For blue and yellow, a number of studies have pointed out that these hues fall close to the daylight locus (e.g., variations between the sky and direct sunlight), and thus the tuning of the blue-yellow axis is matched to a dominant source of variation in the color environment [9]. In fact, this idea has been taken further to suggest that the special perceptual nature of some hues lies entirely in the environment. In opponent theory, blue and yellow (and red and green) are pure or unique hues because they arise from the pure isolated responses within the blue-yellow or red-green channels. However, a problem for this theory is that neurons with the chromatic tuning predicted by these sensations (e.g., that respond only to pure blue-yellow signals) have yet to be discovered. This has suggested the possibility that observers may learn to perceive the unique hues as special because they are special in the environment and not because they correspond to a unique pattern of activity within the neurons coding color. The number of basic color terms (color words that have high agreement and do not refer to specific objects) and the locations in color space that they denote have also been predicted from analyses of how colors are clustered in images. By this account, red spectra tend to form a distinct cluster of signals, and this salient cluster may be why a prevalent color term for red arose in many of the world's languages. More recent analyses of scene statistics have found that objects (at least as defined by image databases) tend to be warm colors while backgrounds tend to be cool colors, and these differences may also

predict the emergence of different linguistic color categories [10].

Regardless of their neural basis, it is clear that like gray, the unique hues are not strongly tied to the specific spectral sensitivity of the observer. The wavelength that appears pure yellow shows little dependence on the relative numbers of long- and medium-wavelength-sensitive cones, even though these can vary over an enormous range of ratios. Similarly, the wavelengths chosen as representing unique hues cannot be predicted from individual differences in lens or macular pigment density and like gray show little effect of the observer's age. These results could again suggest that the hues are set more by the environment. However, it is important to note that the stimuli perceived as unique hues also vary widely across observers and thus cannot be tightly bound to pronounced and stable properties of the world [11]. Indeed, it remains debated to what extent and in what ways color appearance and color naming reflect properties of the observer, the physical environment, or the cultural environment.

A further potential tie of blue and yellow to the environment is that, in a number of tasks, observers appear less sensitive to this color axis. For example, when individuals adjust a stimulus to appear white, they are much more tolerant of blue-yellow variations than red-green variations, and differences in the achromatic settings between observers also exhibit greater variance in blue and yellow. Similarly, color discrimination thresholds are (in some but not all cases) higher for blue and yellow compared to the sensitivity for other chromatic axes. At suprathreshold, blue-yellow variations can have lower effective contrast (e.g., in a color search task), are rated as lower in visual discomfort (which correlates with perceived contrast), and have recently been found to induce weaker cortical responses as measured by functional MRI. One potential explanation for this bias against blue-yellow is that the world itself has a bias toward more blue and yellow. In particular, in many natural outdoor scenes, the variance in chromaticity is higher along bluish-yellowish axes [12]. The visual system adapts not only to the



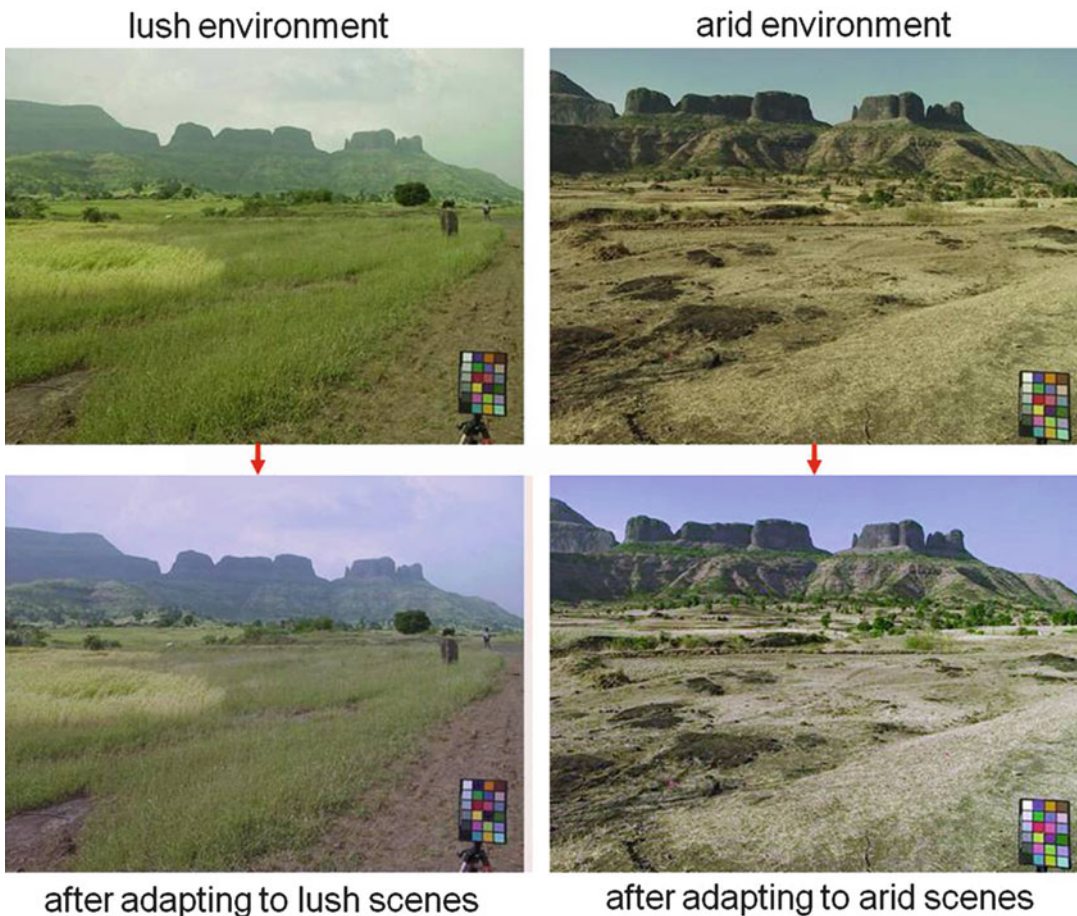
Environmental Influences on Color Vision, Fig. 1 Uniform color systems and natural image statistics. *Left:* Munsell colors plotted within a color space defined by the two color-opponent axes coded by the retinal: long-versus medium-wavelength cone signals (0–180 deg axis) and short versus medium and long cone signals (90–270 deg axis). Sets of points with the same shading represent contours of constant chroma. These are elongated in the bluish-yellowish direction, consistent with weaker

sensitivity to these hue directions. *Right:* color distribution characteristic of natural outdoor scenes plotted in the same space. The scenes have a blue-yellow bias corresponding to the sky and earth. Note that the scaling is chosen to equate sensitivity to each axis, and this roughly matches the range of color along either axis. The weaker sensitivity to blue and yellow may be an adaptation to the greater blue-yellow variance in environmental color

average color in the stimulus but also to the range or contrast. Adaptation to the higher blue-yellow contrasts in typical outdoor scenes may thus leave us less sensitive to these hues. Notably, this bias is also built into the structure of perceptually uniform color spaces (Fig. 1). Color order systems like the Munsell system or spaces like the CIE uniform chromaticity diagram are also stretched along bluish-yellowish axes (though these tend to be closer to an orange-cyan axis). Not surprisingly then, uniform spaces may mirror the structure of the color distributions we are adapted to and in

turn might hold clues to the particular color environment we are calibrated for. Recently for example, a new uniform color metric was proposed based on environmental color statistics and how these should be optimally encoded, and this theoretically derived space approximates color similarities about as well as conventional spaces based instead on purely empirical measures of color discrimination [13].

Finally, environmental factors have also been postulated as the source of perceptual differences between blue and yellow. These differences were



Environmental Influences on Color Vision, Fig. 2 Simulations of how the world might appear to observers adapted to different environments. *Top*: images of the same setting in wet (*left*) or dry (*right*) seasons. *Bottom*: the same images with the colors adjusted to simulate the effects of visual adaptation to the average color and to the range of colors in the different settings. Note this

shifts the average color toward gray and tends to highlight the more novel colors in the environment (Reproduced with permission from Juricevic, I. and Webster, M.A. (2009). Variations in normal color vision. V. Simulations of adaptation to natural color environments. *Visual Neuroscience* 26, 133–145)

brought to the fore by the image of #thedress which different observers perceived as blue-black or white-gold, with the explanation that the differences reflected whether the dress was perceived to be in direct light or shadow. Shadows are more likely to be blue and observers tend to discount the blue in shadows and objects and show greater color constancy for blue variations in lighting [14].

Variations in the Color Environment

Clearly, the world is not stationary in its color characteristics and can instead vary widely from one location to the next or in the same place over time as the lighting or seasons change [15]. If observers are adapted to the colors they are immersed in, then we should expect variations in the properties of their color vision as we go from forest to desert or from summer to winter (Fig. 2). The effects of such environmental changes have not been widely studied, but there are telling signs that they occur. For example, recent studies have pointed to seasonal variations in color mechanisms, from photopigment opsin expression to color appearance to color preferences. Variations in visual sensitivity have also been postulated to reflect differences across environments in exposure to the phototoxic effects of sunlight.

A further important source of environmental variation is the artificial visual worlds that humans increasingly inhabit because of culture and technology. As a result the visual diet of an individual can vary in large and idiosyncratic ways depending on their interests and professions. This could in turn lead to very different color percepts. Neitz et al. exposed subjects for several hours at a time to “red” or “green” worlds by altering their room lighting or by having the subjects wear tinted contacts [16]. This led to persistent changes in the wavelength that appeared unique yellow, consistent with a long-term renormalization in color vision. New lighting technology also alters the gamut of colors, and it remains to be seen how and in what ways color vision will adapt to these environmental changes

[17]. However, to the extent that we understand the processes of adaptation, and can estimate the color distributions that observers are exposed to, it may be possible to predict how a color environment will look to someone adapted to it. These simulations can help reveal very long-term consequences of different color environments for perception, and the extent to which individuals living in different environments might have different or similar color experiences [18].

Cross-References

- [Adaptation](#)
- [Color Contrast Adaptation](#)
- [Color Scene Statistics, Chromatic Scene Statistics](#)
- [Daylight Illuminants](#)
- [Evolution of Primate Color Vision](#)
- [Photoreceptors, Color Vision](#)
- [Unique Hues](#)

References

1. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001)
2. Maloney, L.T.: Physics-based approaches to modeling surface color perception. In: Gegenfurtner, K., Sharpe, L. (eds.) *Color Vision: from Genes to Perception*, pp. 387–416. Cambridge University Press, Cambridge (1999)
3. MacLeod, D.I.A.: Colour discrimination, colour constancy, and natural scene statistics (the Verriest lecture). In: Mollon, J., Pokorny, J., Knoblauch, K. (eds.) *Normal and Defective Colour Vision*. Oxford University Press, London (2003)
4. Mollon, J.D.: Tho’ she kneel’d in that place where they grew... the uses and origins of primate colour vision. *J. Exp. Biol.* **146**, 1–38 (1989)
5. Tregillus, K.E., Engel, S.A.: Long-term adaptation to color. *Curr. Opin. Behav. Sci.* **30**, 116–121 (2019)
6. Webster, M.A.: Adaptation and visual coding. *J. Vis.* **11**(5), 3, 1–3, 23 (2011)
7. Werner, J.S., Scheffrin, B.E.: Loci of achromatic points throughout the life span. *J. Opt. Soc. Am. A.* **10**, 1509–1516 (1993)
8. Delahunt, P.B., Webster, M.A., Ma, L., Werner, J.S.: Long-term renormalization of chromatic mechanisms following cataract surgery. *Vis. Neurosci.* **21**, 301–307 (2004)

9. Shepard, R.N.: The perceptual organization of colors: an adaptation to regularities of the terrestrial world? In: Jerome, H., Cosmides, L., Tooby, J. (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 495–532. Oxford University Press, New York (1992)
10. Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Conway, B.R.: Color naming across languages reflects color use. *Proc. Natl. Acad. Sci.* **114**(40), 10785–10790 (2017)
11. Rolf, G.K.: Variability in unique hue selection: a surprising phenomenon. *Color Res. Appl.* **29**(2), 158–162 (2004)
12. Webster, M.A., Mollon, J.D.: Adaptation and the color statistics of natural images. *Vis. Res.* **37**, 3283–3298 (1997)
13. Smet, K.A.G., Webster, M.A., Whitehead, L.A.: A simple principled approach for modeling and understanding uniform color metrics. *JOSA A*. **33**(3), A319–A331 (2016)
14. Witzel, C., Gegenfurtner, K.R.: Color perception: objects, constancy, and categories. *Annu. Rev. Vis. Sci.* **4**, 475–499 (2018)
15. Webster, M.A., Mizokami, Y., Webster, S.M.: Seasonal variations in the color statistics of natural images. *Netw. Comput. Neural Syst.* **18**(3), 213–233 (2007)
16. Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., Williams, D.R.: Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron*. **35**, 783–792 (2002)
17. David, A., Smet, K.A.G., Whitehead, L.: Methods for assessing quantity and quality of illumination. *Annu. Rev. Vis. Sci.* **5**, 479–502 (2019)
18. Webster, M.A.: Probing the functions of contextual modulation by adapting images rather than observers. *Vis. Res.* **104**, 68–79 (2014)

ERP

► Visual Evoked Potentials

Event Related Potential

► Visual Evoked Potentials

Evolution

► Environmental Influences on Color Vision

Evolution of Color Expression

David Alan Warburton

Institute for the History of Ancient Civilizations,
Northeast Normal University, Changchun, China

Synonyms

Materiality, luxury, hue, saturation, brightness, intensity linguistically expressed in ancient languages; Problems of the earliest language development in the domain of color; Understanding perception and expression of color in ancient languages

Definition

Conceptual issues related to perception and expression in color categorization, relevant to understanding the impact of ancient thought on the development of abstraction.

Introduction

For Grand-Clément (in [1]) “The ancient Greeks considered golden the best color in the realm of metals and purple as the best color for textiles.” In Greek, *porphúreos*, modern “purple” was an expensive red dye, associated with royalty and divinity, and the sea, and she therefore suggests that ancient Greek expressions of color relied on

economic as well as symbolic values [expressed using] substances [, with] gold and purple [being prominent examples,...] substances could be endowed with specific values and highly esteemed [...] we have to keep in mind that the ancient Greek perception of colors differs from that of the modern viewer: the Greeks [...] paid more attention to brightness and vividness than hues. Moreover, they perceived colors not as abstract categories but rather as visual qualities of objects and substances. This conception of color, deeply rooted in the material world can be observed in the terminology, since adjectives such as *khruuseos* [= *xrúseos*, “golden”], *argureos* [“silver, silver-plated”], *khalkos* [= *xálkeos*, “bronze, brazen”] indicated both the

physical matter and the color of an object. Thus, the first question should be: what were the colors of gold and purple from the Greek perspective?

She is largely correct, but, inadvertently abbreviates and distorts the larger picture. On the one hand, (a) the texts and art offer only the expression and not the perception. On the other, as she starts with Greek material, she fails to stress (b) that both “gold” and “purple” are loan-words in Greek, and thus also neglects (c) the Near Eastern influences that began the conceptual appreciation of color, without abstraction. The Near Eastern use of the names of stones and metals to designate artificial imitations began a tradition leading to the designation of hues. Greek developments are highly significant, but merely a segment of the era of continuous interaction, during which loan-words acted as cognitive levers, lifting meaning.

It was a long era in terms of individual human experience, lasting around five millennia – but insignificant in terms of the biological cognitive capacities of the modern human mind, which probably go back 300,000 years or so. The beginning of abstract color terminology was a cognitive process taking place relatively rapidly in relatively recent times, relying on materials. By the end of the fourth millennium BC, the Egyptian and Mesopotamian elites had become accustomed to the shiny, malleable metals gold and silver along with exotic hard pure lapis lazuli and carnelian. Ultimately, it was from this encounter that conceptual yellow, white, dark blue and deep red eventually emerged, but it was a long process, so that even among the Greeks – thousands of years later – conceptions of color still involved luxury, hue, intensity, brightness, hardness, purity and form (see also Encyclopedia of Color Science and Technology entries on ► [“Materials and the Origins of Abstract Color Naming”](#) and ► [“Ancient Color Terminology”](#)).

“Purity” itself was probably conceptually, linguistically and culturally defined for the first time in this era through refined, pure, “white” silver in Mesopotamia. By the second millennium BC, this approach led to multiple “reds” and “blues”; the high priced precious metals and stones having embodied and enhanced the understanding of

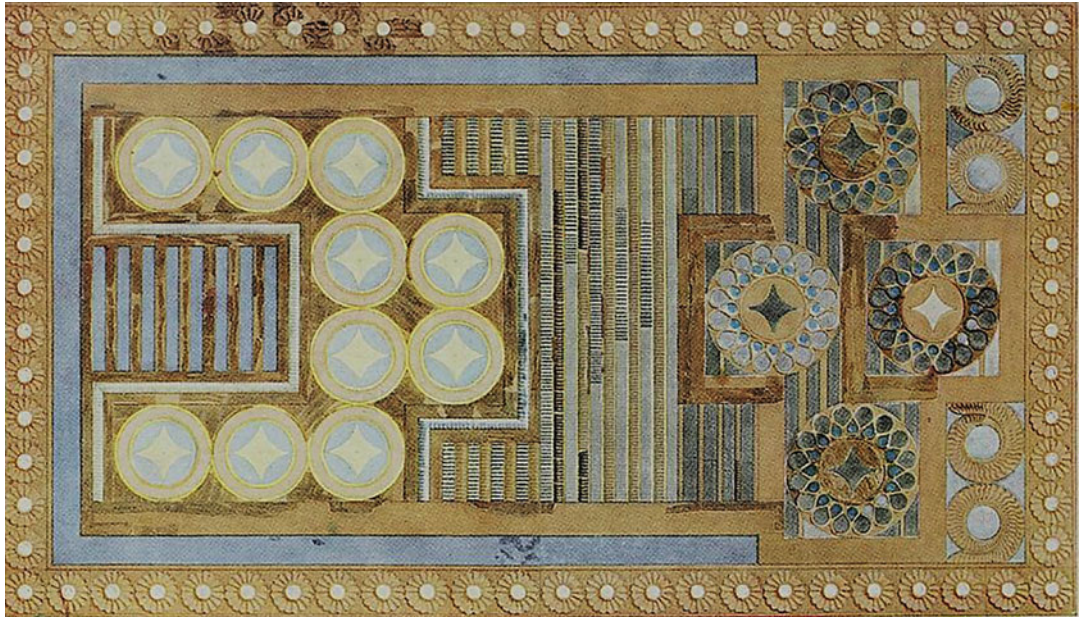
purity and truth expressed through color during the third millennium BC, the formative period for Bronze Age thought.

In (i) the evolution of color expression and perception, the Classical Greeks (in contrast to the Homeric Greeks!) were far closer to a modern understanding of color than the Egyptians and Mesopotamians millennia earlier, as (ii) loan-words (e.g., Greek *xrusós* for gold and *kuáneos* for “blue” were related to the Akkadian *hurāšum* and *uqnû* respectively), and (iii) foreign concepts – silver paved the way for the Mesopotamian invention of the concept of “purity” in “white” which spread across the Old World. Extraordinary is that the range of meaning in the Greek terms *argēeis* (derived from silver) and *leukós* (abstract white) are virtually identical in “shining,” etc., with *leukós* (sic) alone also meaning “pure”; this allowed complex abstract color concepts to develop figurative meanings.

Grand-Clément says that “Aphrodite, the goddess of erotic desire and sexual attraction, was “golden” (*khruśeē*) [= *xruśeē*; and like the irresistible attraction of gold itself,] her radiance was so powerful and extreme that no one – human or divine – could escape her seductive power”; the same powers were attributed to Mesopotamian Inanna/Ishtar and Egyptian Hathor – but only Hathor was the “woman’s gold.” (Gold was a late comer in the Mesoamerican world, yet Mayan and Aztec “yellow” is also associated with “want,” “desire,” and “mature” [3]).

Ereshkigal (the Mesopotamian Queen of the Netherworld) lives in chambers of lapis lazuli, but the apartments of her Greek counterpart Persephone are “blue” or “dark” [4] – using the Greek color term *kuáneos*, derived from the Akkadian designation of lapis lazuli *uqnû* (cf. Fig. 1). But the Greeks knew only the color concept, and the Mesopotamians only the material.

Yet Gladstone [5] concluded that the Homeric Greeks could not see blue – being at a lower stage of human evolution (anticipating Darwin’s *Origins of Species*, published the following year). Today, this is biological nonsense: visual perception of color has not changed in the last 300,000 years. But the difficulty of recognizing blue and green in Ancient Greek linguistic



Evolution of Color Expression, Fig. 1 Inlaid gaming board (96.5 × 55.3 cm) – of silver, ivory, gold, rock crystal, faience, ground glass paste – from Knossos, Crete, ca. mid-second millennium BC [2]. Note both yellow = gold and contrast of light and dark blue; Evans himself [2] uses the word “kyanos” (= *kíanos*) for the

cobalt blue glass paste (of Egyptian origin) used to make the dark blue color (*kuáneos*) imitating lapis lazuli. Craftsmen were handling the materials for the benefit of the elite who became accustomed to color being associated with exotic beautiful things – and this influenced the poets

expression remains – and differs from the concrete appreciation of blue already a millennium before Classical Greek (Figs. 1 and 2). Abstraction in language arrived as a cognitive consequence of what was begun by human contact with the esteemed materials [9, 10]. Blue lost out to purple, as Greek purple was linked to expensive dye while Greek blue was linked to cheaper glass (and not the real lapis lazuli).

The Evolution of the Linguistic Division of the Spectrum

Empirically, first millennium BC Greek and Chinese terminology is not that different from second and third millennium BC Sumerian, Egyptian, and Akkadian. Conceptually, however, in the first millennium BC, both Greek and Chinese conceived of color differently, with a tendency to increasing adjectival abstraction, but decreasing clarity about hue.

The earliest use of red ocher by modern humans dates back more than 100,000 years.

Along with black soot, red and yellow ocher set the limit on color in the Upper Paleolithic cave paintings ca. 35,000 years ago (Moutsiou in [11]). Green and blue are absent until tens of thousands of years later. Green stones obtained prominence beginning approximately 13,000 years ago (Bar-Yosef Mayer in [11]). Red carnelian, black obsidian, blue lapis lazuli, white limestone, blue-green turquoise, jade, gold, and silver gradually and progressively become prominent during the following millennia: all present by the fourth millennium when writing was invented. Until the first millennium BC, color designations remained limited, employing materials as proxies.

Grasping the transformation is difficult, but recognizable. The later Chinese word for “red” *hóng*, had a (possibly linguistically related) homonym meaning “rainbow.” For the Chinese, the rainbow is associated with the celestial brightly colored dragon [12]. The second millennium BC Akkadian understanding of the rainbow (*tirana*) was associated with the form of the weapon



Evolution of Color Expression, Fig. 2 Restored interpretation of so-called “Partridge Frieze” (detail, approximately 15×40 cm) at Knossos, Crete, mid-second millennium BC [6]. The craftsmen were not only handling precious materials characterized by color, but also deliberately creating paints to depict the world – and using color

as the only means available in flat art. Note distinction of different blues and reds, browns (perhaps slightly less than a millennium before Homer). For technical aspects of painting as a means of expression in Antiquity, cf., e.g., [7, 8]

(^{gis}*tir*) – and the same applies to the Egyptian *pdj.t*, designating both weapon and meteorological phenomenon – but not color. This distinction may identify the moment – after the middle of the first millennium BC – when abstract colors appeared in human language and thought. Visible in the Chinese language, it was appearing in Greek thought; the Ancient Near East never reached this level. Significantly, the Aztecs had a completely detailed chromatic understanding of the rainbow (Dupey-García in [1]) – and this seems to build on Greek and Chinese conceptual developments.

The sequence was not built on adding additional abstract words for additional perceived colors, but rather one whereby individual parts of the spectrum were split in accord with materials before adding further abstract color terms – and the difficulties proved insurmountable in ancient Near Eastern thought. Abstraction accelerated when the Greeks and Chinese overcame the Ancient Near Eastern system of exhaustive lists as a means of classifying knowledge, reducing the world to handful of “elements” or “phases,” including colors (Bogushevskaya, Warburton in [1]).

At the moment, there are four dominant theoretical approaches to understanding the emergence of color vocabulary in the earliest

languages. Consciously or unconsciously, many authors – and especially those working on Maya [3] – maintain the framework of the Berlin & Kay Theory. While relying on materiality, Maya seems to build on developments in the Greek and Chinese worlds – leading to a chromatic paradigm roughly compatible with the Berlin & Kay Theory (for which support from MacLaury’s Mesoamerican color survey was fundamental). The recent history of color may be compatible with the Berlin & Kay approach – but accounts neither for origins nor does it reflect the original acquisition sequence.

Since before Berlin & Kay, working with Egyptian and earlier theoretical models, Schenkel (in [11, 13, 14]) started with hue oppositions, but has concentrated on abstract words, defined by (a) verbal (rather than adjectival) categories, demonstrating (b) that a word for “grue” never appeared in Egyptian, as the focus of the relevant word was in green, identifying (c) a failure of historical salience for “white,” and revealing (d) that the fifth color term in Egyptian was a second “red” (rather than a “yellow”). Although superficially similar, Schenkel’s theoretically and empirically founded theoretical system differs fundamentally from Berlin & Kay, while more restricted (but dealing with the most ancient continuous language dataset available).

Working on a different theoretical basis with Akkadian, Thavapalan [15] follows current linguistic and anthropological trends stressing luminosity and saturation more than hue in the formation of the earliest color categories. However, the terminology she identifies demonstrates that materials (stones, glass, metals, dyes, etc.) were decisive for the lexical expression of color.

Working exclusively with developments in ancient color usage, Warburton (in [1, 7, 9, 11, 16]) stresses the importance of (a) the language-based evidence of the precious materials (strong hues in rare hard materials) for color in (b) providing the basis for the emergence of the earliest advanced categorization of color, (c1) enabling the subsequent course of the development of color terminology through loan words, and (c2) guiding the use of the names of the minerals for the artificial products (e.g., lapis lazuli for glass, turquoise for faience), with hue and material going together: material was the foundation whereby abstract terminology was established in the course of recent millennia – and the “materiality” approach is gaining traction (contributions in [1, 11]).

Warburton concurs with Schenkel that the possibilities of the terms for precious materials were not exploited in the Bronze Age so that real abstraction came later in Greece and China. Thavapalan stresses saturation and brightness, but both Thavapalan and Warburton recognize the key role of materials as decisive in the expression of color in the earliest documented languages, and they thus move further along the spectrum than Schenkel who sees a preliminary level of abstraction necessarily excluding “blue” and “yellow.” “Materiality” is gradually becoming the dominant approach to understanding the origins of color expression in language, but for Schenkel (in [11]), this remains otherworldly.

(See also Encyclopedia of Color Science and Technology entries on ► [“Ancient Color Terminology,”](#) ► [“Materials and the Origins of Abstract Color Naming,”](#) ► [“Berlin and Kay Theory,”](#) and ► [“Mesoamerican Color Survey Digital Archive”](#)).

Further Complications for Applying Modern Standards to Ancient Materials

Plato considered terms for “bright,” “brilliant” (e.g., *lamprón*, *stilbon*) to designate colors [17], i.e., identifying “colors” where modern approaches seek hues, with “brightness” an attribute. Ancient Chinese also already had homonyms (e.g., *mìng*), which meant each individually implied some nuance of “brightness” – with “dark” another problem.

The Pre-Greek color terminology was more complicated; e.g., the ancient Egyptian word *ṯn.t* for “faience” also meant “dazzling,” but probably originated in a designation for a type of pale green natural glass (Corcoran, Warburton in [1]). Words like English “black” and “white” are different from “light,” “shining,” “gleaming,” “sparkling,” “dark,” and “gloomy,” respectively, as the latter glosses appear to not have category foci. (See also Encyclopedia of Color Science and Technology entries on ► [“Ancient Color Terminology”](#) and ► [“Materials and the Origins of Abstract Color Naming”](#)).

This is a major obstacle to the application of the Berlin & Kay theory as it implicitly suggests that “dark”/“black” and “bright”/“white” are the same – whereas they are not: artistically (like the Minoans, Fig. 3), the Egyptians went to great lengths to achieve pure blacks and whites in their art (e.g., Warburton in [7]; also Blom-Böer in [11]) – and these were conceptually salient, even if the range of hue categories incorporated into any given word could be broad in the eyes of observers like Schenkel (in [11]).

In early languages the word for a generic “red” may have served as a synonym for “colorful” or “colored.” Peculiarly, the – very late! – Aztec *tlalpalli* (“red”) played a similar role (Dupey-García in [1]). While the color typically labeled “red” in modern English language usage was among the earliest distinguished in art and written artifacts, the concept of a category of a single “red” as a distinct linguistic unit was not dominant in the second and third millennia BC: “red” was distributed among several different words in various languages.

Although often considered “abstract,” the adjectives for colors in Sumerian were never



Evolution of Color Expression, Fig. 3 Restored interpretation of a Mycenaean dagger (detail of blade L: 20 cm) from Greece, mid-second millennium BC [18]. It is Minoan craftsmanship. The bronze blade has been treated so as to accommodate a band of niello plating to form the

shiny black background surface, into which copper, “polished silver,” and “pale gold” inlays were set. The effect of the brightness is due to the pure black background from which the copper, silver and gold shine forth. Pure bright colors were the goal

attached to the *NAM* abstract determinative (in contrast to Chinese where the *sè* radical is added today). Given the difficulties of understanding adjectives in Egyptian, Schenkel used verbs, i.e., *ḥdj* “to be white,” *km* “to be black,” *dšr* “to be red,” and *wādj* “to be green”; although verbs, they served as adjectives; the noun-based *xsbḏjy* “lapis lazuli-like” “dark blue” was a metaphor, simile, or adjective; likewise *mḥkātj* “turquoise-like” for “light blue” or “light-green,” and *nbwj* for “gold,” “golden.” Following Schenkel, the adjectival aspect was introduced through the precious materials.

Early on, loan-words were crucial, as Akkadian *uqnū* and Egyptian *xsbḏj* for “lapis lazuli” or “dark blue” are two different loanwords for an imported material, borrowed from other unknown languages. Other terms shared in different languages named materials but not “abstractions.”

Detailed Linguistic Issues in the Debate About the Significance of Ancient Color Naming

Berlin & Kay’s theory pre-supposed a teleological, universalist, evolutionary convergence in the modern linguistic partition of the spectrum with languages progressively adding additional abstract terms. Baines’s [19] application of the

Berlin & Kay sequence to the earliest languages has been recognized to have failed (e.g., [16], Corcoran in [1]); the empirical support for the acquisition sequence is missing; lack of salience and the importance of loan words throw further doubt on its overall validity. (See also Encyclopedia of Color Science and Technology entry on ► “Berlin and Kay Theory”).

Black and White, Dark and Light

Significantly, a word for “darkness” shared in Sumerian (*gege*, *kuku*) and Egyptian (*kk.w*) means that the word was of great antiquity, since it fed into two distinct language families – or the Sumerian influenced the Egyptian as an early loan word. Yet these glosses for “darkness” were used only in Sumerian to influence “black,” but not in Egyptian where the origin of “black” lay elsewhere (Egyptian *km* for “black” may be related to soot). The original and later Chinese terms for “black” (*xuán*, *hēi*) did mean “dark” – but the Chinese term for white *bái* did not serve as a major designation for “bright” or “clear” (more *míng* and *qīng*). According to the Berlin & Kay framework, there is an implied assumption that “dark” & “black” and “bright & white” can be equated (implying color words which are not necessarily present).

Another peculiarity is clear evidence, in both earlier Greek and earlier Egyptian, of a tendency

for a “red” and “white” opposition [13, 19, 20]; archeological evidence would imply that a tripartite opposition of red, black and white can be traced back to around 11,000 years ago, and thus many millennia before the verbal oppositions of these same three colors are confirmed in written sources. The first conceptually isolated color category might have been neither black nor white, but rather red; the distinction of “black” and “white” (possibly recognized earlier only as “light” and “dark”) may have been the consequence of metallurgy generating “purity,” so that white silver was conceptually distinguished after the cognitive recognition of “red” as a general category, different from “light” and “dark.” The division of “red” into several hues was the first step in the lexical partition of the spectrum, with the categorization of “black” and “white” only taking place over the last 5000 years, with white probably playing the role of cognitive catalyst, being the result of creating “purity” = silver = “white” – and leading to the identification of soot = refuse? = “black” as the opposite of purity = “white.” The black and white opposition can only be understood in terms of purity, and thus followed from earlier developments.

Salience, Convergence, Usage and Meaning

Chinese *bái* (“white”) is documented as an early color word (in Chinese – but more than a millennium after the earliest Ancient Near Eastern texts), largely focused in the color recognized today – and maintained throughout history, but *bái* is also associated with silver. In its final stage, the early Egyptian for “white,” *ḥdj*, is replaced by *ūbaš* (earlier meaning “to shine”). The salient meaning of *ḥdj* in Egyptian is the material silver, not the color word. The concept of “white” = “purity” = “silver” seems to be prevalent, and probably came from Mesopotamia.

In Chinese, the contemporary word for black, *hēi*, replaced *xuán*, the word for black used during the first millennium BC; evidence of historical salience for “red” in Chinese is likewise lacking, as the ancient term *chì* was abandoned and substituted with *hóng*. In Mycenaean Greek, the main word for “red” is *po-ni-ko-ro* (later *phonikós*, a loanword referring to the Phoenicians

who furnished the red dye) rather than *e-ru-to* (which gave rise to the Classical basic color term *eruthrós*).

Classical Greek *kuáneos* for blue still exists in Modern Greek, but has ceded dominance to *mple* = “blue,” a loan word – and thus blue is not salient, and the modern word a recent loan word. Materials and loanwords are more evident than salience, but the progression underlying the Berlin & Kay approach is based on independent abstract terms accumulating gradually and independently, and remaining stable. A rather uniform terminology has arisen – but through exchange of materials, loan-words and contact. The teleological approach works when projecting from the present back, but not when beginning with the origins and moving forward.

For Egyptian, Akkadian, and Sumerian (dating to before the middle of the second millennium BC), the color terminology did not match the colors used in art. In contrast to this, already for Mycenaean Greek and Classical Greek (thus, from the second half of the second millennium BC onward), “it is hardly surprising that in the use of color, language and representative art are two sides of the same coin” [20]. Thus, in comparing Bronze Age paintings with Roman-era painting, for Blakolmer, the use of color in painting reflects the development of a vocabulary in the Greek language. By comparison, in Egyptian art, there are many colors used in painting that are not accounted for by any kind of color-related vocabulary (Warburton in [7], also Blom-Böer in [11]).

The salience emerging from the first millennium onward depends on a fundamental change during the second millennium; this linguistic and cognitive change influenced later developments.

Color as Categorization

This discussion aims at the origins and early development of color terminology, rather than use. In general, philologists catch the color terms where they appear and not because the ancients systematically used colors as an elementary means of classification. Nor did the ancients

view gold, silver, lapis lazuli, turquoise, carnelian, jade, amethyst, and the like as ordinary products, but rather as extraordinary and unusual materials: the bones, flesh, and hair of the gods would be silver, gold, and lapis lazuli in ancient Egypt. Classical Greek divine statues are described as *xruselephantēlektros*, “of gold, ivory, and electrum”: attributes of divinity, not mere high-priced materials, but part of another world (stressed of Egyptian color conceptions by Schenkel in [11]). The process of creating color categorization involved demystifying and banalizing precious materials, rendering them cognitively flexible – betraying the divine origins.

Summary

The dominance of the Berlin & Kay Theory has blocked access to the understanding of how color terminology developed during the earliest stages. Precious materials and loanwords played a decisive role in forming the foundations for developing an early understanding of color beyond the simplest terms among the literate. In the earliest phase, the variety of terms for “red,” “green,” “yellow,” and “blue” probably preceded the expansion of color vocabulary beyond Berlin & Kay Stages III and IV. Account must be taken of the origins and development of early color terminology; recent convergence may likewise owe more to exchange and loanwords than to universalism. The use of ordinary materials to designate color is the final form of the process practiced after abstraction had come into existence. Developing abstractions was largely a matter of hue.

Cross-References

- [Color Categorical Perception](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effects of Color Terms on Color Perception and Cognition](#)
- [World Color Survey](#)

References

1. Goldman, R. (ed.): *Essays in Global Color History: Interpreting the Ancient Spectrum*. Gorgias Press, Piscataway (2016)
2. Evans, A.: *The Palace of Minos at Knossos*, vol. I. Macmillan, London (1921)
3. Houston, S.D., Brittenham, C., Mesick, C., Tokovinine, A., Wariner, C.: *Veiled Brightness: A History of Ancient Maya Color*. University of Texas Press, Austin (2009)
4. Liddell, H.G., Scott, R., Jones, H.S.: *A Greek-English Lexicon*. Oxford University Press, Oxford (1958)
5. Gladstone, W.E.: *Studies on Homer and the Homeric Age*. Oxford University Press, Oxford (1858)
6. Evans, A.: *The Palace of Minos at Knossos*, vol. II, part I. Macmillan, London (1928)
7. Brinkmann, V., Primavesi, O., Hollein, M.: *Circumlitio: The Polychromy of Antique and Mediaeval Sculpture*. Schriftenreihe der Liebieghaus Skulpturensammlung, Frankfurt/Main (2010)
8. Davies, W.V. (ed.): *Colour and Painting in Ancient Egypt*. British Museum Press, London (2001)
9. Warburton, D.A.: Colourful meaning: terminology, abstraction, and the near eastern Bronze Age. In: Johannsen, N., Jessen, M., Jensen, H.J. (eds.) *Excavating the Mind: Cross-sections Through Culture, Cognition and Materiality*, pp. 183–208. Aarhus University Press, Aarhus (2012)
10. Warburton, D.A.: The importance of the origins of abstraction and discourse. In: Bredholt Christensen, L., Jensen, J.T. (eds.) *Religion and Material Culture*, pp. 61–99. Brepols, Turnhout (2017)
11. Thavapalan, S., Warburton, D.A. (eds.): *The Value of Color: Material and Economic Aspects in the Ancient World Berlin Studies of the Ancient World*, vol. 70, Berlin (2019). <https://doi.org/10.17171/3-70>
12. Schuessler, A.: *ABC Etymological Dictionary of Old Chinese*. University of Hawai'i Press, Honolulu (2007)
13. Schenkel, W.: Die Farben in Ägyptischer Kunst und Sprache. *Z. Ägyptische Sprache*. **88**, 131–147 (1963)
14. Schenkel, W.: Color terms in ancient Egyptian and Coptic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 211–228. Benjamins, Amsterdam (2007)
15. Thavapalan, S.: *The Meaning of Color in Ancient Mesopotamia Culture & History of the Ancient Near East*. Leiden, Berlin (2019)
16. Warburton, D.A.: The theoretical implications of ancient Egyptian colour vocabulary for anthropological and cognitive theory. *Lingua Aegyptia*. **16**, 213–259 (2008)
17. Ierodiakonou, K.: Plato's theory of colours in the *Timaeus*. *Rhizai*. **2**, 219–233 (2005)
18. Evans, A.: *The Palace of Minos at Knossos*, vol. III. Macmillan, London (1930)

19. Baines, J.: Color terminology and color classification: ancient Egyptian color terminology and polychromy. *Am. Anthropol.* **87**, 240–262 (1985)
20. Blakolmer, F.: Zum Charakter der frühägäischen Farben: Linear B und Homer. In: Blakolmer, F. (ed.) *Österreichische Forschungen zur Ägäischen Bronzezeit 1998: Akten der Tagung am Institut für Klassische Archäologie der Universität Wien 2.–3. Mai 1998*, pp. 225–239. *Wiener Forschungen zur Archäologie*, Vienna (2000)

Evolution of Primate Color Vision

Romann M. Weber¹ and Mark Changizi²

¹DisneyResearch|Studios, Zurich, Switzerland

²2ai Labs, Institute for the Study of Human and Machine Cognition, Columbus, OH, USA

Synonyms

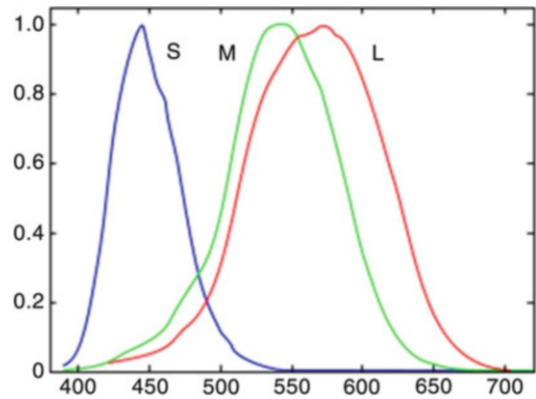
[Cone fundamentals](#); [Dichromacy](#); [Primate color vision](#); [Trichromacy](#)

Definition

Comparative color categories are salient partitions in the perceived color spaces of primates that are known to vary when encoded by signaling systems that differ across species.

The Neurophysiology of Color Vision

The entire range of human color perception, in all its vibrancy, is due to the operation of just three types of photosensitive retinal cone cells and the neural mechanisms responsible for interpreting their signals. These cone cells contain photosensitive opsin pigments that preferentially absorb light at different wavelengths. The short-wavelength (S) pigment has peak absorption of light at about 430 nm, roughly corresponding to blue light; the medium-wavelength (M) pigment has peak absorption at about 530 nm, corresponding to green light; and the long-wavelength (L) pigment has peak



Evolution of Primate Color Vision, Fig. 1 Sensitivity curves for human photopigments (attribution: Vanessaekowitz at en.wikipedia)

absorption at about 560 nm, corresponding to red light (see Fig. 1).

Dysfunction in the cone cells or in the coding of their pigments leads to various types of color blindness. In humans, genes for S-type cones are located on chromosome 7, while those for the M- and L-type cones are located on the X chromosome. This largely explains why color blindness is a much more common phenomenon in men, as women possess independent copies of the M- and L-pigment genes on their two X chromosomes.

The M- and L-type pigments are nearly identical in their genetic specifications. The small difference in their light absorption profiles is due to a difference in only 3 out of 364 amino acids that code for their respective proteins [8]. The S-type pigment, on the other hand, is more distinct on the molecular level.

Each type of cone cell responds over a range of wavelengths, and the intensity of a cone's response to a light signal depends not only on that signal's wavelength but also on its intensity. This creates a situation in which the wavelength of a light signal cannot be uniquely identified based on the response of a single type of cone cell. Unique color information is only recoverable through the simultaneous and concerted action of different types of cone cells operating within a small neighborhood of each other.

In a sense, the response profile of each type of cone cell can be viewed as a “basis vector” that,

along with the other cones that make up the basis, defines a multidimensional color space. This analogy is made more concrete when considering the *color-matching* experiment that is foundational to the field of color science. In this experiment, observers are presented with a test light with an arbitrary spectral power distribution on one side of a bipartite white screen. The observers' task is then to try to match the appearance of the test light by individually adjusting the intensities of a series of primary lights focused on the other side of the screen. When a match is achieved, the lights on each side of the screen, while physically different in their spectral distributions, are perceived to be the same. Appearance-matched lights produced in this way are called *metamers*. There is a degree of variability among color-normal observers in the amounts of the three lights required to achieve a metameric match.

For people with normal color vision, exactly three primary lights are required to perform the color-matching task. As a result, normal human color vision is said to be three-dimensional or *trichromatic*. Establishing the dimensionality of color vision in nonhuman animals is an often difficult task requiring extensive training and testing of the animals [7]. A frequent practice in the literature – arguably justified – is to equate an animal's color vision dimensionality with the number of distinct color photopigments it expresses in its retina. However, the tarsier provides an example of a primate with a fairly uneven retinal distribution of its two cone types, challenging classical views about how color identification works through local networks of opponent cone cells [5]. The tarsier's labeling as a dichromat based on a photopigment-type count alone should therefore be done cautiously [7].

The Evolution of Primate Color Vision

It is believed that gene duplication and mutation of cone pigments, which ultimately made multidimensional perceptual color spaces possible, traces back to the time of the first vertebrates [1]. Almost all vertebrates have an S-type pigment very similar to that possessed by humans [8].

Most mammals have a single X-chromosome cone opsin gene, typically coding for a single long-wavelength visual pigment, and are *dichromats*. But even among primates, only the Old World monkeys, apes, and humans are *trichromats*, having a second – and highly homologous – cone opsin gene on the X chromosome, which codes for a medium-wavelength visual pigment [8, 12]. The generation of these red- and green-pigment genes is thought to have occurred following a duplication event after the split between the New World monkeys and the *catarrhines*, a group of higher primates made up of the Old World monkeys and apes (including humans), about 40 million years ago [12].

Opsin-gene and cone-pigment complements appear to be largely the same among the catarrhines [8, 12]. Psychophysical studies of macaques provided much early information on the color discrimination abilities of the Old World monkeys [7], and much of this research showed performance very similar to that of humans, suggesting a concordant perceptual color space [15]. Studies of chimpanzees also showed wavelength discrimination similar to human performance [4] (for a review, see [6, 7]).

Among the placental mammals, trichromacy appears to be found only in certain primates [7]. However, it is possible to enhance color discrimination behavior in dichromatic mammals through genetic engineering. Behavioral tests have shown strong evidence for novel color vision in mice genetically engineered to express the human L-opsin gene [9]. This suggests that the visual systems of traditionally dichromatic mammals can readily make use of higher-dimensional color signals.

Why, then, when traditionally dichromatic animals can be easily made to demonstrate three-dimensional color vision, is trichromacy seemingly restricted to certain primates? Further, why does trichromacy in primates manifest itself the way that it does, namely, with such a large spectral overlap between the M- and L-pigment sensitivities? Several hypotheses have been advanced to answer these questions. They are of two basic types: one focusing on the potential advantage of three-dimensional color vision in finding food and

the other suggesting that trichromatic color vision was selected for perceiving subtle, behaviorally relevant cues on exposed skin.

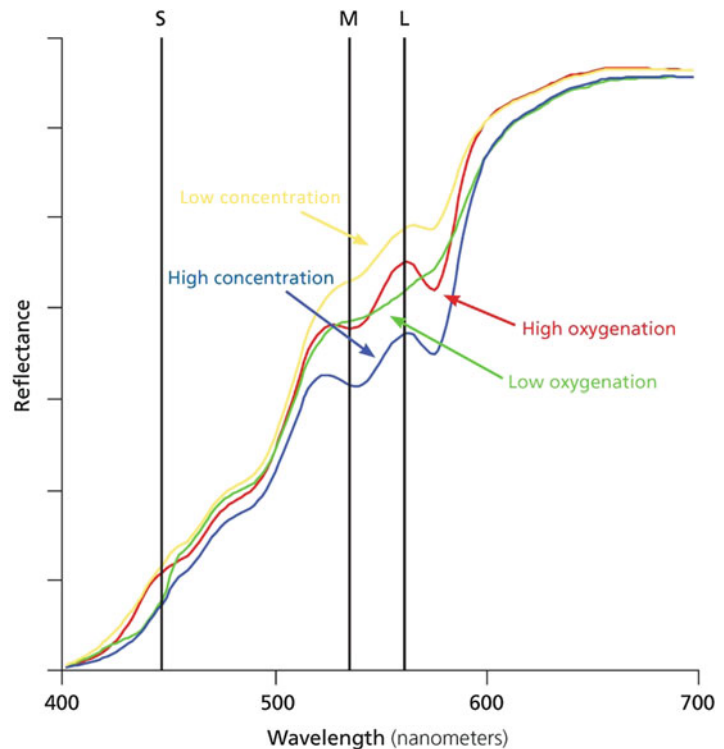
The experience of humans with various forms of reduced color vision suggested the hypothesis that color vision was selected for its benefit to detecting targets in visual scenes. Specifically, it has been conjectured that color vision is especially useful for detecting ripe fruit and edible leaves [11]. Consistent with this hypothesis is the observation that ripe fruit is more easily detected against a background of leaves in a trichromatic perceptual space [13]. Also, in one study, several species of frugivorous (fruit-eating) primates tended to eat and disperse seeds from fruit whose reflectance spectra are tightly clustered in the color space described by trichromatic catarrhine photopigments, while non-preferred fruit is more diffusely distributed in this color space [14]. In another study, folivorous (leaf-eating) trichromatic primates that were surveyed tended to eat leaves shifted more toward the red end of the spectrum relative to leaves eaten by non-trichromatic primates [10].

The food-based hypothesis suggests an obvious evolutionary advantage for primates. However, a question remains as to why catarrhine trichromacy is so uniform in its cone sensitivities given the wide ranges of diets – not all of them strictly frugivorous or folivorous – of these primates. An alternative hypothesis suggests that the specific tuning of primate color vision is selected for its advantage in detecting changes in the concentration and oxygenation of blood in conspecifics as indicated by subtle color shifts displayed on exposed skin [3].

Skin reflectance can be modulated over two dimensions according to oxygen saturation and concentration of hemoglobin in the skin [2, 3]. The effect of changes over the oxygenation-concentration dimensions results in shifts in skin reflectance that are predictable (see Fig. 2), with oxygenation modulations shifting hues across the red-green axis and concentration modulations shifting hues across the blue-yellow axis.

Particularly notable in this pattern of reflectance changes is that they occur exclusively in the region of color space at which the M and

Evolution of Primate Color Vision, Fig. 2 Skin reflectance changes due to varying oxygenation and concentration of blood



L pigments are maximally sensitive. That is, the characteristic “W” shape nestled within skin’s reflectance profile shifts its position in the region of peak sensitivity of the most recently evolved photopigments, which are nearly optimal for detecting this shift [3]. These shifts are also consistent across all skin types.

It has been observed that trichromatic primates tend to have either bare faces or large areas of exposed skin, such as a bare rump, a finding supportive of the skin-based hypothesis [3]. Further, dichromats are demonstrably poor at detecting changes in skin coloration [2]. Also, whereas the reflectance spectra of the collective primate diet vary widely, the shifts in skin reflectance across the oxygenation-concentration dimensions are stable across all primates [3].

More general shifts of skin color are often linked with various conditions of health and emotional status [3]. The skin-based hypothesis for trichromatic vision selection is unique in that it suggests a *social* advantage for catarrhine color vision. This advantage applies not only to maintaining the integrity of the social group through a sensitivity to health-related cues but also to perceiving threats from conspecifics, such as signs of aggression in the form of increased oxygenation of blood.

Cross-References

- [Color Categorical Perception](#)
- [Cone Fundamentals](#)
- [Environmental Influences on Color Vision](#)
- [Metamerism](#)
- [Trichromacy](#)

References

1. Allman, J.M.: *Evolving Brains*. Scientific American Library, New York (2000)
2. Changizi, M.A.: *The Vision Revolution*. BenBella Books, Dallas (2009)
3. Changizi, M.A., Zhang, Q., Shimojo, S.: Bare skin, blood and the evolution of colour vision. *Biol. Lett.* **2**(2), 217–221 (2006)

4. Grether, W.F.: Chimpanzee color vision. I. Hue discrimination at 3 spectral points. *J. Comp. Psychol.* **29**, 167–177 (1940)
5. Hendrickson, A., Djajadi, H.R., Nakamura, L., Possin, D.E., Sajuthi, D.: Nocturnal tarsier retina has both short and long/medium-wavelength cones in an unusual topography. *J. Comp. Neurol.* **424**, 718–730 (2000)
6. Jacobs, G.H.: *Comparative Color Vision*. Academic, New York (1981)
7. Jacobs, G.H.: Primate color vision: a comparative perspective. *Vis. Neurosci.* **25**, 619–633 (2008)
8. Jacobs, G.H., Nathans, J.: The evolution of primate color vision. *Sci. Am.* **300**, 56–63 (2009)
9. Jacobs, G.H., Williams, G.A., Cahill, H., Nathans, J.: Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science*. **315**, 1723–1725 (2007)
10. Lucas, P.W., et al.: Evolution and function of routine trichromatic vision in primates. *Evolution*. **57**, 2636–2643 (2003)
11. Mollon, J.D.: Tho she kneel’d in that place where they grew. . . . *J. Exp. Biol.* **146**, 21–38 (1989)
12. Nathans, J., Thomas, D., Hogness, D.S.: Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science*. **232**, 193–202 (1986)
13. Osorio, D., Vorobyev, M.: Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. B.* **263**, 593–599 (1996)
14. Regan, B.C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., Mollon, J.D.: Fruits, foliage and the evolution of primate colour vision. *Philos. Trans. R. Soc. B.* **356**, 229–283 (2001)
15. Sandell, J.H., Gross, C.G., Bornstein, M.H.: Color categories in macaques. *J. Comp. Physiol. Psychol.* **93**(3), 626–635 (1979)

Expressive Rendering

- [Non-photorealistic Rendering](#)

Extended Dynamic Range

- [High Dynamic Range Imaging](#)

Eye Sensitivity Functions

- [Spectral Luminous Efficiency](#)

F

Factors of Color Categorization

► [Mechanisms of Color Categorization](#)

Favorite Colors

► [Color Preference](#)

Fechner, Gustav Theodor

Mark D. Fairchild
College of Science, Rochester Institute of
Technology, Rochester, NY, USA



Gustav Theodor Fechner was a German experimental psychologist and philosopher. He is also considered by many to be the father of modern psychophysics. Initially, Fechner took a degree in medicine and worked in that area for a while. During that time he began publishing a series of humorous and satirical articles and poems lampooning the medical profession. These were published under the pseudonym Dr. Mises, and one of his most famous publications of this genre was on the *Comparative Anatomy of Angels* (1825). The writings of Dr. Mises certainly provide insight into the breadth and depth of Fechner's intelligence and abilities.

Fechner then moved on to physics by learning about contemporary advances in electricity and magnetism through translation of great French works as well as handbooks of chemistry and physics into German. With this new knowledge in hand, Fechner taught at the University of Leipzig, eventually obtaining a professorship and making distinguished contributions to the field. Interestingly, Dr. Mises also continued to publish on occasion.

All along, Fechner was interested in the mind-body problem and desired to determine an empirical relationship between the physics of the world (and body) and the perceptions of the mind. Some of his studies in this vein included the discovery of perceived colors for slowly flickering black-and-white patterns

(known as Fechner's Colors or Fechner-Benham Colors after Benham made the work accessible in English) and the detailed study of color afterimages. It was the study of afterimages that set Fechner off in his next direction. He was studying afterimages by staring at the sun through colored filters. This led him to give up his chair in physics in 1840 due to induced photophobia from eye injuries that made him an invalid, and overly sensitive to light, for about a decade.

During this period, on October 22, 1850, while lying in bed he finally figured out the basis of linking physical measurements in the environment with human perceptions in the mind that is currently referred to as Fechner's law. Based on his knowledge of Weber's law (coined by Fechner) that, for many perceptions, the ratio of a just-noticeable change in a stimulus to the initial magnitude of a stimulus is a constant, Fechner figured out that the differential equation implied by Fechner could be integrated and assumed that the just-noticeable differences could be summed to predict perceptual magnitude. The resulting relationship, now termed Fechner's law, mathematically suggested that the magnitude of a perception would be proportional to the logarithm of the stimulus intensity.

More modern knowledge tells us that the specific relationship depends on the perceptual quantity and that it is not valid to sum JNDs to predict magnitudes. Nonetheless, Fechner's contribution was very significant in founding the field of psychophysics. Fechner published further details of the theory and practice of psychophysics in his seminal work, *Elements of Psychophysics* (1860), [1] which remains a useful guide for practitioners in the field. Interestingly, Fechner is also credited with being the first to introduce the concept of the median to formal data analysis.

From Boring's introduction in Adler's translation of *Elements of Psychophysics*, we find that Fechner was for 7 years a physiologist (1817–1824), for 15 years a physicist (1824–1839), for 12 years an invalid (1839–1851), for 14 years a psychophysicist (1851–1865), for 11 years an experimental estheticist (1865–1876), and for periods throughout a philosopher.

References

1. G. Fechner: *Elements of Psychophysics* (trans.: H. E. Adler). Holt, Rinehart and Winston, New York (1966)

Fechner's Colors and Benham's Top

John Krantz

Natural Sciences, Department of Psychology,
Hanover College, Hanover, IN, USA

Synonyms

[Benham's disc](#); [Fechner-Benham subjective color](#); [Illusory colors](#); [Pattern induced flicker colors](#); [Subjective colors](#)

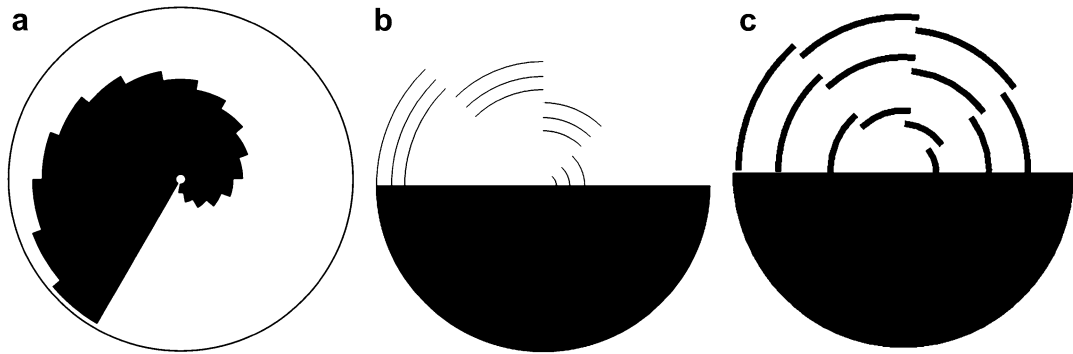
Definitions

Fechner colors refers to the illusory colors that result from the repeated flashing of black and white patterns. These colors are often referred to as Fechner colors who is often given credit for their discovery.

Benham's top is a pattern developed for a spinning disc to create illusory colors that have become a standard.

Basic Phenomenon and History

Fechner colors refers to the creation of illusory or subjective colors through an alternating pattern of black and white stimulus. The first report of these subjective colors was by Prevost, a French monk [5]. He used a white light and put a white sheet of paper to break the beam periodically to generate the illusory colors. Fechner is often reported as the discoverer of this illusion and his name has been given to these illusory colors. Fechner's creation of the subjective colors was accidental as he was at the time was involved in the creation of discs to create different shades of gray [1, 5] (Fig. 1a). Fechner was the first to use a spinning disc as



Fechner's Colors and Benham's Top, Fig. 1 Discs used to create subjective colors. (a) One used by Fechner, (b) one used by Benham, (c) a common one based on Benham's work

F

the means for creating these illusory colors. There were several studies throughout the nineteenth century of this phenomenon with spinning discs including by such a luminary as Helmholtz. Helmholtz even proposed the early explanations for these colors. Each of the early studies had used individually designed discs to create the illusory colors. In 1894, Benham developed what has become the near standard disc pattern of producing these illusory colors that he called the “artificial spectrum top” and has come to bear his name [2]. This disc is called usually either Benham's top or Benham's disc. Figure 1 shows some examples of the discs that have been used, including the one by Fechner (Fig. 1a), one by Benham (Fig. 1b), and another common one derived from Benham's (Fig. 1c). Variations in the disc include having thicker single lines or thinner three parallel lines.

To generate these illusory colors, it is important that the flicker rate of the stimulus is not too slow or too fast. The flicker is noticeable and the stimulus is well short of the critical fusion frequency. The rate of alternation of the black and white areas is not terribly high with effective stimuli frequencies going from 3 Hz on up with optimal rotational frequencies for the spinning disk of about 4–6 Hz [3, 4]. The direction of the spinning seems to effect the colors experienced [11].

Creating and Measuring Fechner Colors

To create Fechner colors, the disc is usually spun in the frequencies reported above. Colors are

experienced in bands located where the black lines will be flickering. While Benham's top represents a traditional means of creating the colors, researchers have developed many alternatives in trying to understand how these illusory colors occur [4, 8, 10, 11]. It is also not a necessity that a spinning disc or top be used. A stationary flickering stimulus can be used as [8] though there is some evidence that the colors are not as easily seen [10]. To see an online example that uses a flickering stimulus the reader is referred to Krantz [9].

Measuring the colors produced by Benham's top has been approached several ways over the years. The most common methods use some form of color matching to determine the color experienced by the participant. Schramme [12] used a color matching paradigm where a mirror was used to overlay the color patch onto the spinning disc. Then participants tried to match this patch to the illusory colors they were seeing. Rosenblum, Anderson, and Purple [10] used Munsell color chips to make the color matches which served them well as they were interested if the illusory colors show some of the features of color matching errors seen in dichromats. Jarvis [8] used a binocular matching method where the illusory colors were presented in one eye and the matching color was presented in another eye. In all of these methods, the attempt was to try to determine the color perceived by the observer.

Results from various studies over the years have revealed several interesting features of Fechner colors. First, the colors experienced are always desaturated, that is somewhat washed out

[8]. These colors can vary in their vibrancy in ways that depends upon the method of generation, but natural colors can much more easily be made to be highly saturated. Many reports indicate a wide range of individual differences in the colors experienced [7]. While most visual phenomena have some individual variation, there is a remarkable degree of agreement across observers. If there were not this agreement, then technologies such as color reproduction would not be feasible. This agreement is not seen in Fechner colors. Other researchers have observed that both illumination level and adaptation state alter the colors experienced, making the adaptation state of observers important to control in any experiment on Fechner colors [3, 4]. Perhaps one of the most interesting findings is that Fechner colors can be obtained with monochromatic, single wavelength, illumination as well as in full-spectrum neutral illumination [3].

Explanations and Future Directions

There have been many different types of explanations offered for the existence of Fechner colors in the almost 200 years since its discovery [6]. The challenge for any theoretical explanation is twofold: how does a monochromatic or neutral color stimulation can lead to the perception of colors, and what is the role of the flickering rate. Another important factor to consider, often ignored in most theoretical attempts, is the individual differences in the colors reported by observers. Older explanations have involved fatigue in receptors, the role of complementary colors, and even factors related to Hering's original theory of color perception. Most of these explanations have been ruled out by directly examining fatigue, contrast, and the limitations of the complementarity of the colors experienced [1]. As more information about cone functioning developed, explanations developed that used ideas about differences in the speed with which different wavelengths of light are processed [3] and information about the relative speed with which the blue or short-wavelength cone responds to flickering stimuli compared to the speed of the other cones respond to the same

flickering stimuli have been developed. The hypothesis is that it is the relative slowness of the short-wavelength cone in responding to flickering stimuli that is responsible for Fechner colors. The idea is that with slowly flickering stimuli, the short-wavelength cone responds too slowly to pick up the flicker adequately while the middle and long wavelength cones, red and green cones, respectively, still can respond. Thus, these cones do not adequately cancel each other out in the blue-yellow color opponent channel leading to the perception of colors. Schramme [11] found, using color matches, Fechner colors that fell along the blue-yellow axis of the 1931 CIE diagram that matched these expectation.

Despite the sophistication and intuitive plausibility of this explanation, it is still safe to say that there is no generally accepted explanation for these illusory colors. Campenhausen et al. [4] propose a two-step model for Fechner colors that involves both summing of cone inputs in a non-color opponent pathway and lateral interactions at the level of the horizontal cells. The support for this hypothesis comes from many lines of evidence but includes that Fechner color in areas that do not receive modulated light and that achromatic effects similar to Fechner colors can be produced in rod vision. Schramme [11], whose results seemed to support the possibility that Fechner colors were due to the slowness of the short-wavelength cone, proposed that Fechner colors actually arise in the blue/yellow ganglion cells. The argument here depended upon the ability to manipulate the perceived Fechner color along the blue-yellow axis of the CIE diagram in a manner would have been hard to predict from the action of the short-wavelength cone alone. Using these two explanations as examples, it can be seen that even the level of the visual system involved is not agreed upon, though all theories do seem to argue that Fechner colors arise in the retina.

While these explanations are intriguing, future developments will be needed to combine, modify, or seek a new direction of explanation for Fechner colors. One feature of the findings that is rarely explained and needs to be is the relative range of individual variation in the colors experienced. The actual mechanism of the temporal response

needed to be further uncovered as this finding increases the evidence for multiple ways that the eye responds to temporally moving or flickering stimuli. It is possible that it is in the variations in temporal response that the individual differences of Fechner colors lie.

Cross-References

- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [Color Order Systems](#)
- [Color Phenomenology](#)
- [Cone Fundamentals, Stockman-Sharpe](#)
- [Motion and Color Cognition](#)
- [Psychological Color Space and Color Terms](#)

References

1. Bagley, F.W.: An investigation of Fechner's colors. *Am. J. Psychol.* **13**(4), 488–525 (1902)
2. Benham, C.E.: The artificial spectrum top. *Nature.* **51**, 200 (1894)
3. Brown, J.L.: Flicker and intermittent stimulation. In: Graham, C.H. (ed.) *Vision and Visual Perception*, pp. 251–320. Wiley, New York (1965)
4. Campenhausen, C., Hostetter, K., Schramme, J., Tritsch, M.F.: Color induction vs non-opponent lateral interactions in the human retina. *Vis. Res.* **32**(5), 913–923 (1992)
5. Cohen, J., Gordon, D.A.: The Prevost-Fechner-Benham subjective colors. *Psychol. Bull.* **46**(2), 97–136 (1949)
6. Falk, D., Brill, D., Stork, D.: *Seeing the Light*. Harper & Row, New York (1986)
7. Gerjuoy, H., Clarke, F.R.: Fechner colors on television. *Am. J. Psychol.* **71**(3), 606–607 (1958)
8. Jarvis, J.R.: On Fechner-Benham subjective color. *Vis. Res.* **17**, 445–471 (1977)
9. Krantz, J.H.: Benham-Fechner Colors and Behnam's Top. <https://isle.hanover.edu/Ch06Color/Ch06FechnerColors.html> (2015). Accessed 18 June 2019
10. Rosenblum, K., Anderson, M.L., Purple, R.L.: Normal and color defective perceptions of Fechner-Benham colors: implications for color vision theory. *Vis. Res.* **21**, 1483–1490 (1981)
11. Schramme, J.: Changes in pattern induced flicker colors are mediated by the blue-yellow opponent process. *Vis. Res.* **32**(11), 2129–2134 (1992)

Fechner-Benham Subjective Color

- [Fechner's Colors and Benham's Top](#)

Femto-Photography

- [Transient Imaging](#)

Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics

Vladimir G. Chigrinov

Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong
Department of Theoretical Physics, Moscow Region State University, Mytishi, Russia
Nanjing Jingcui Optical Technology Co., Ltd, Nanjing, China

Synonyms

[Displays](#); [Ferroelectric liquid crystals](#); [Field sequential](#); [Photoalignment](#); [Photonics](#)

Definition

A ferroelectric liquid crystal (FLC) is one in which the molecules that constitute the material are chiral, possess a dipole moment in the direction which is perpendicular to the long molecular axis, and are tilted with respect to the formed layers. This is characteristic of the smectic C* liquid crystal. They have application in a number of photonic and display devices.

Applications

Fast tunable photonic elements are one of the hot research topics these days. Thus, liquid crystals have become attractive for applications such as diffractive optics, adaptive optics, or optical metrology [1]. While nematic liquid crystal (LC) devices are the dominant technology, because of its easy manufacturing, devices based on ferroelectric liquid crystals (FLCs) have grown in importance because they can switch up to 100 times faster than standard nematic LCs at the cost of less power consumption [2]. FLCs have found applications in diffractive optics for the generation of digital holograms, adaptive beam steering systems, and polarization gratings. Additionally, high-speed communication systems and producing elements (such as switches, attenuators, and polarization rotators) demand extremely fast response, stable, low loss, operable over a wide temperature range, and that require small operating voltages and extremely low power consumption. Thus, fast switching ferroelectric liquid crystal (FLC) is a good candidate for the new generation of photonic and display devices, which proved to be better in response time, than usually used nematic LC. However, the FLCs are less popular for real applications because of difficult alignment process [3–10].

The photoalignment technology enables the key problems usually faced in FLCD applications to be solved [11], such as (i) quality of FLC alignment on sufficiently large surface area; (ii) appropriate adjustable anchoring energy and pretilt angle; and (iii) low losses in the alignment layers due to their small thickness. The further development of novel photoaligned fast ferroelectric LCD and photonics devices aimed at (i) fundamental study of the new appropriate electrooptical modes used for switching such as electrically suppressed helix (ESH) FLC and deformed helix FLC (DHFLC); (ii) understanding of physical mechanisms of FLC interaction with a photoaligned surface of different photosensitive nature (diffusion, photocrosslinking, and photo-degradation) to produce a stable alignment with a controllable anchoring energy and pretilt angle over sufficiently large

surface area; (iii) development of new fast response FLC materials with fast switching and low power consumption; (iv) elaboration of fast response optical FLC elements for photonics applications, such as switches, grating, Fresnel lens, Circular Dammann (CD) grating, attenuators, and polarization rotators; and (vi) investigations of regimes of operation to allow to use efficient addressing of FLCD.

The response time of conventional LC devices is still inferior to the existent competing technologies. The target LC cells' response time for various photonic and display devices is time < 1 ms to provide a high resolution, low power consumption, and extended color gamut LC displays. Fast switching FLC is a good candidate for the new generation of liquid crystal display and photonics devices, which proved to be much better in response time than usually used nematic LC. The problems, which are usually faced in FLCs such as (i) quality of FLC alignment on sufficiently large surface area; (ii) appropriate adjustable anchoring energy and pretilt angle; and (iii) low losses in the alignment layers due to their high thickness of the alignment layer, etc., can be effectively solved by a photoalignment technology. Multistable electrooptical modes were shown to exist under certain conditions in photoaligned FLC cells, which mean that any light-transmission level can be memorized after the driving voltage is switched off. High-frequency hysteretic free electrically controlled $0-2\pi$ phase modulation of the light can be implemented using a very short helix pitch (less than 200 nm) DHFLC. The application of fast V-shaped DHFLC for new active-matrix LCD and optical data processing devices is envisaged. On the other hand, the ESHFLC mode manifests saturated electro-optical up to frequency 5 kHz at the electric field of 1 V/ μm .

At the same time, application of FLC for high-speed communication systems, enables elements (such as switches, attenuators, and polarization rotators) that are extremely fast, stable, durable, of low loss, and operable over a wide temperature range, and require small operating voltages and extremely low power consumption when using bistable and multistable switching

with optical memory. Passive FLC elements for fiber optical communication systems (DWDM components) based on LC cells can successfully compete with the other elements used for the purpose, such as micro electromechanical, thermo-optical, opto-mechanical, or acousto-optical devices.

The prototypes of the novel photoaligned FLC photonic elements may include high resolution, low power consumption, and high efficiency, which can be used for the beam-steering devices, lens, microdisplays, and other adaptive optical elements. The switchable lenses based on new FLC prototypes can be efficiently applied in the new generations of switchable 2D/3D LCD TV. Thus, developed fast and low power-consuming FLCs can also be deployed for the Field Sequential Color (FSC) FLC microdisplay, which is now one of the most advanced technologies, for pico-projectors can be also made on the basis of new materials and electrooptical modes in FLC to be developed in the project. The modulators and switches based on fast responded FLC are envisaged in Dense Wavelength Division Multiplexing (DWDM) elements of optical communication systems, such as switches, filters, attenuators, equalizers, etc.

Recent Developments

An extensive work studying optical and electro-optical response of FLC, as well as developing new FLC materials with advanced surface alignment, has been conducted. A good progress in formulating new photoalignment materials that are potentially capable of being used in photo-aligning technology for LCDs was made. The aligning materials developed at HKUST are based on photo-polymerized and cross-linked dye photosensitive layers that enable the following: (i) high order parameter >0.8 ; (ii) excellent alignment quality of LCD with a high contrast ratio in TN, VAN, π -BTN, and FLC modes; (iii) high temperature stability up to 230 °C suitable for LCD manufacturing; (iv) UV-stability due to the polymerization and cross-linking effect in dye

layers; (v) perfect adhesion and high azimuthal and polar anchoring energy, i.e., $>10^{-4}$ J/m², which is the same as the anchoring of the rubbed polyimide (PI) layer; (vi) excellent sensitivity with a minimum energy of 150 mJ/cm² for a nonpolarized obliquely incident light and 20 mJ/cm² for polarized light, allowing roll to roll processing of retardation films and polarizers; and (vii) the value of voltage holding ratio (VHR) for the photo-aligned LC cell ($>99\%$ at 80 °C) and residual DC voltage (<50 mV) was found to be even better than those for rubbed PI layers.

Fast switching FLC with V-shape switching, with a number of gray scales, which can be used, on the basis of field sequential colors (FSC) FLC LCD have been developed. Several FLC electro-optical modes for the matter have been employed, including electrically suppressed helix (ESH), deformed helix ferroelectric (DHF), and vertical aligned DHF (Kerr effect) modes. Novel photo-aligned FLC devices may include FSC FLC with a high resolution, low power consumption, and extended color gamut, which can be used in the screens of portable PCs, mobile phones, and PDAs. The switchable goggles and lenses based on new FLC prototypes can be efficiently applied in the new generations of switchable 2D/3D LCD TV. FSC FLC microdisplay, which is now one of the most advanced technologies for pico-projectors, can be also made on the basis of new materials and electrooptical modes in FLC.

Fast ferroelectric LC (FLC) are applied for high speed communication systems, producing elements (such as switches, attenuators, and polarization rotators) that are extremely fast, stable, durable, of low loss, and operable over a wide temperature range, and that require small operating voltages and extremely low power consumption when using bistable and multi-stable switching with optical memory. Passive FLC elements for fiber optical communication systems (DWDM components) based on liquid crystal (LC) cells can successfully compete with the other elements used for the purpose, such as micro electromechanical (MEM), thermo-optical, opto-mechanical, or acousto-optical devices.

References

1. Chigrinov, V.G.: Liquid Crystal Photonics, 165 pp. Nova Science Publishers (2014)
2. Guo, Q., Yan, K., Chigrinov, V., Zhao, H., Tribelsky, M.: Ferroelectric liquid crystals: physics and applications. *Crystals*. **9**(470), 1–21 (2019)
3. Shi, L., Srivastava, A.K., Tam, A.M., Chigrinov, V.G., Kwok, H.S.: 2D–3D switchable display based on a passive polymeric lenticular lens array and electrically suppressed ferroelectric liquid crystal. *Opt. Lett.* **42**(17), 3435–3438 (2017)
4. Shi, L.Y., Srivastava, A.K., Cheung, A., Hsieh, C.T., Hung, C.L., Lin, C.H., Lin, C.H., Sugiura, N., Kuo, C.W., Chigrinov, V.G., Kwok, H.S.: Active matrix field sequential color electrically suppressed helix ferroelectric liquid crystal for high resolution displays. *J. Soc. Inf. Disp.* **26**(5), 325–332 (2018)
5. Budaszewski, D., Srivastava, A.K., Chigrinov, V.G., Woliński, T.R.: Electro-optical properties of photo-aligned photonic ferroelectric liquid crystal fibres. *Liq. Cryst.* **46**(2), 272–280 (2019)
6. Ma, Y., Tam, A.M.W., Gan, X.T., Shi, L.Y., Srivastava, A.K., Chigrinov, V.G., Kwok, H.S., Zhao, J.L.: Fast switching ferroelectric liquid crystal Pancharatnam-Berry lens. *Opt. Express*. **27**(7), 10079–10086 (2019)
7. Guo, Q., Xu, L., Sun, J., Yang, X., Liu, H., Yan, K., Zhao, H., Chigrinov, V.G., Kwok, H.S.: Fast switching beam steering based on ferroelectric liquid crystal phase shutter and polarisation grating. *Liq. Cryst.* **46**(9), 1383–1388 (2019)
8. Srivastava, A.K., Pozhidaev, E.P., Chigrinov, V.G., Kwok, H.S.: Vertically aligned ferroelectric liquid crystals with high Kerr constant for field sequential color displays. *J. Mol. Liq.* **295**, 111054, 1–5 (2019)
9. Liu, Y., Chen, P., Yuan, R., Ma, C.-Q., Guo, Q., Duan, W., Chigrinov, V.G., Hu, W., Lu, Y.-Q.: Ferroelectric liquid crystal mediated fast switchable orbital angular momentum of light. *Opt. Express*. **27**(25), 36903–36910 (2019)
10. Guo, Q., Liu, T., Wang, X., Zheng, Z., Kudreyko, A., Zhao, H., Chigrinov, V., Kwok, H.-S.: Ferroelectric liquid crystals for fast switchable circular Dammann grating. *Chin. Opt. Lett.* **18**(8), 080002 (104) (2020)
11. Chigrinov, V., Guo, Q., Kudreyko, A.: Photo-aligned ferroelectric liquid crystal devices with novel electro-optic characteristics. *Crystals*. **10**, 563 (2020). <https://doi.org/10.3390/cryst10070563>

Ferroelectric Liquid Crystals

► [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

Field Sequential

► [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

Filling-in

► [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Finish, Textile

Parikshit Goswami

School of Design, University of Leeds, Leeds, West Yorkshire, UK

Synonyms

[Dry or mechanical finishing](#); [Textile processing](#); [Textile wet processing](#); [Wet or chemical finishes](#)

Definition

Textile finishing can be defined as all processes (chemical and/or mechanical), employed subsequent to textile coloration which impart additional functionality/superior aesthetics to the textile material. Mostly, textile finishing is applied to fabrics (woven, knitted, nonwovens); however, textile finishes can also be applied to fibers and yarns.

Introduction

The primary objectives of textile finishing are to improve the aesthetic and functional properties of textiles. In a broader sense, “finishing” relates to all processes that fabrics might be subjected to subsequent to weaving, knitting, or nonwoven manufacturing processes. The term “finishing”

could include fabric preparation (e.g., singeing, desizing, scouring, bleaching, optical brightening process, mercerization process, etc.) and coloration (dyeing and printing); indeed, the combination of these processes is sometimes referred to as “wet-finishing” processes [1]. Another school of thought refers to “finishing” as the final stage of fabric preparation with the objective to prepare the fabric for the consumer and that the term “finishing” concerns only fabric treatment other than fabric preparation and coloration. In this chapter, finishing subsequent to coloration will be discussed, in which context, it is important to check the compatibility of the finishing process with the treated substrate and the coloration process that has been previously implemented [2].

Classification

The conventional textile finishing methods can broadly be divided into two categories:

- (i) “Dry” or “mechanical” finishes
- (ii) “Wet” or “chemical” finishes

There are, however, finishing operations which combine mechanical and chemical finishes e.g., mercerization (the NaOH treatment of fabric on machines with or without tension).

“Dry” or “Mechanical” Finishes

The mechanical finishing of textiles may range from a simple drying operation to a complicated series of calendaring operations (Schreiner finisher). A few examples of the mechanical finishing of textile are discussed below.

Drying

It is well known that wet fabric can hold a large amount of water and that some of this held water can be easily removed by simple mechanical action. The efficiency of the process to mechanically remove the water from the fabric is vitally important from both cost and environmental points of view. High energy is required to

evaporate the remaining water due to its high latent heat of evaporation. Machines such as mangles, centrifuges, and suction-slot machines are used in textile processing to remove water from fabric by mechanical means [3]. Drying machines are employed in textile processing with the purpose of removing the water that has not been removed by mechanical action. The majority of the dryers used in textile processing offer continuous throughput and a few examples of such machines are “hot flue”, “drum dryers”, “perforated drum dryers”, “tumble dryers”, etc. [1].

Calendering

In calendering, the fabric is passed between two heavy rollers. The rollers may vary in hardness, surface speed, and temperature. Fabric properties such as smoothness and luster can be improved using calendering [4]. Calender finishing can be of six types, namely, simple finishing or swizzing calendering, chasing calendering, friction calendering, Schreiner calendering, embossing calendering, and felt calendering. A few examples of calendering are discussed below.

Friction Calendering

The differential speed of the two rollers is carefully selected to produce effects such as “chintz.” The smooth metal roller normally rotates at a higher rate than the soft roller [1, 3]. Durable finishes (Everglaze) can be achieved using this method by adding a cross-linking agent [5].

Embossed Effects

In this technique, a fabric pattern is produced by calendering the fabric using an engraved heated metal roller and a soft roller [4]. To avoid slippage between the rollers, it is important to control the process accurately.

Schreiner Finishing

In “Schreiner,” fine lines engraved on a metal roller are transferred to the fabric. With appropriate fabric construction and by carefully engineering the line direction of the engraving, a soft lustrous handle can be achieved [3]. This treatment is used to give cotton fabric the appearance of silk [1]. Although the wash fastness of this

effect is poor [1, 3], it increases the point-of-sale appeal [3].

Napping

Napping is a very effective way for imparting a soft handle to fabrics [6]. Fabrics of low-twist, staple fiber yarn can be used for napping [7]. The napping machine contains metal cards and the napping tool comprises rollers with curved metal wires [1] which pull the fiber ends to the surface of the fabric. One or both sides of the fabric can be napped [7].

Sanforizing

Fabrics contain internal tensions as a result of the various manufacturing processes employed, and a pretreatment is required to nullify this effect, since, otherwise, there is a possibility of significant shrinkage of garments during washing. Sanforizing is a controlled mechanical shrinkage process without the use of any chemicals. This treatment should be the last treatment that is applied to the fabric [1].

“Wet” or “Chemical” Finishes

Although chemical finishing has always been an important process for textiles, the demand for high-performance textiles has grown exponentially in recent years and, as a result, so has the demand for chemical finishing [8]. The majority of chemical finishes for textiles are additive, where the finishing treatment results in an increase in substrate mass through sorption or deposition of chemical compounds. However, some chemical finishing can be *subtractive*, in which the treatment results in the fiber substrate losing mass as a result of chemical degradation. While chemical finishing is mainly applied to fabric, it can also be applied to loose fibers, yarns, garments, etc. Based on the life-span of the finish on the textile substrate, textile finishing can also be classified as *transient* and *durable* finishing. The most important aspect in this regard is the resistance of the finish to domestic washing [9].

Application of Chemical Finishes

While the application and formulation of chemical finishes depends on several factors, the most important factors are the composition of the material being treated and the chemistry of the functional chemical. Interaction of the finishing effects, compatibility of different chemicals, and the environmental/green credentials of the process are other factors to consider.

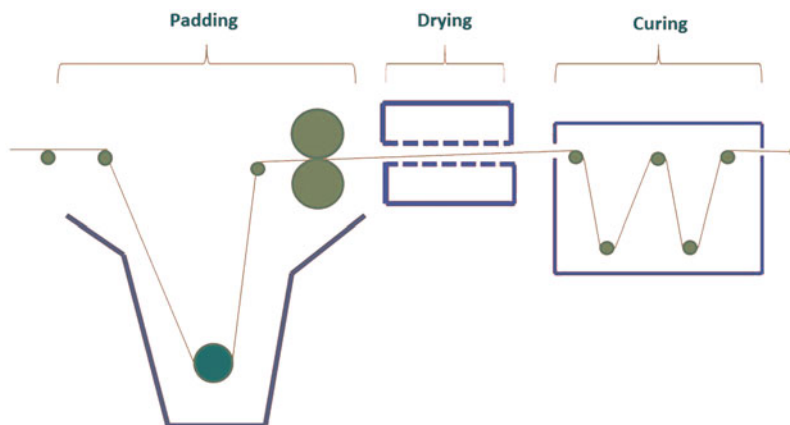
If the functional chemical finishes displays high substantivity towards the substrate, then exhaust (immersion) application methods can be used and any of the textile dyeing machines employed for batchwise dyeing can be used for textile finishing. However, if the applied functional chemical has limited or low substantivity, then a continuous application method can be used. One such method is the pad-dry-cure method wherein the fabric is first immersed in a solution containing the functional chemical, followed by drying and finally “curing” to fix the finish within the fabric (Fig. 1).

Chemical finishes can also be applied to textiles using modern application techniques such as exposure of the textile substrate to plasma employing either low pressure, low-temperature plasma, or atmospheric plasma. Plasma, which is considered as the fourth aggregation state of matter, can be described as a mixture of partially ionized gases in which the constituents are achieved by external energy addition. The main advantage of such plasma treatments are:

- It consumes minimal chemicals compared to conventional finishing.
- No costly drying process is required.
- It is a surface treatment, so the bulk properties of the material are not affected.
- High environmental compatibility.

Important Chemical Finishes

A few commercially important chemical finishes are discussed below. Other types of chemical finishes include, but are not limited to, antistatic finishes, soil release finishes, insect resist finishes, flame-retardant finishes, stain resist finishes, non-slip finishes, UV-protection finishes, finishing to improve color fastness, etc.

Finish, Textile,**Fig. 1** Schematic diagram of pad-dry cure operation

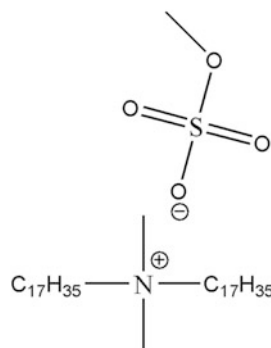
F

Softening Agents

Softeners are one of the most widely used chemicals in textile processing, and as the name suggests, the application of softeners dramatically improves the handle and perceived quality of a fabric. In textile processing, one of the most important functions of softeners is to counteract the harshness imparted by other finishes (e.g., easy care). Softeners are, however, also used extensively in domestic washing formulations.

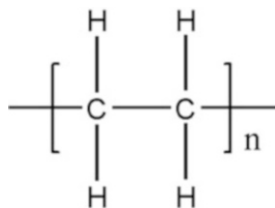
Although the main effects of the softeners are at the surface of fibers, the small molecules can impart internal plasticization of the fibers by reducing the glass transition temperature (T_g) of the fibrous polymer [8]. Early softeners were based on waxes and oils although modern types can be divided into five categories, namely, cationic softeners, nonionic softeners, anionic softeners, reactive softeners, and amphoteric softeners [3].

Cationic softeners are the most important type of softener for both industrial and domestic applications. The positively charged ends of the cationic softeners orient themselves towards the negatively charged fibers (zeta potential), creating a new surface of hydrophobic carbon chains that are responsible for the excellent softening effect of cationic softeners [8]. Because of the positive charge, these softeners are substantive towards cellulosic fibers, and exhaust application of these finishes is widespread using conventional textile dyeing machines (e.g., winch, jet, beam) [3]. Cationic softeners are known to exacerbate the soiling propensity of fibers

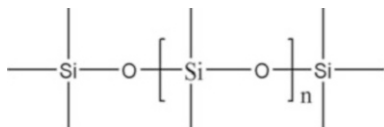
**Finish, Textile, Fig. 2** Dimethyldistearylammonium methosulfate

and to inhibit soil removal. Dimethyldistearylammonium methosulfate is a good example of a popular cationic softener (Fig. 2). The use of silicone-based finishes in textile processing has grown steadily over the last 50 or so years [3]. Silicon-based cationic softeners provide a high degree of softness and a unique hand to fabrics. However, silicon softeners may contain variable amounts of siloxane oligomers, depending on the method of synthesis, which may be a cause of air pollution [8]. A popular range of silicon softeners are based on the weakly cationic amino functional siloxanes. Although these silicon compounds could potentially be applied from mild alkaline conditions, durability is improved by application at pH of 4.

Nonionic softeners are well known to perform multiple functions such as softeners, emulsifiers, stabilizers, extenders, and lubricants [3]. Paraffin waxes and similar materials are the most basic



Finish, Textile, Fig. 3 Polyethylene

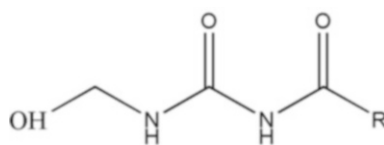


Finish, Textile, Fig. 4 Polydimethylsiloxane

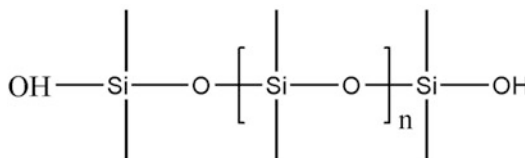
nonionic softeners. Molten polyethylene (Fig. 3) at high pressure can be air oxidized to secure hydrophilic characteristics (mainly carboxyl groups). High-quality, more stable products can be produced by emulsification in the presence of alkali. These products are stable to extreme pH conditions and can withstand the temperatures used for normal textile processing [8]. In recent years, polysiloxanes have gained importance as nonionic softeners (Fig. 4). Due to their limited substantivity, nonionic softeners, with very few exceptions, are normally applied by non-exhaust methods (e.g., padding) [3].

Anionic softeners, due to the presence of the negative charge, are repelled from negatively charged fibers (e.g., cellulosic fibers) which leads to higher hydrophilicity [8]. These were among the first soft finishes to be used commercially and include long-chain alkyl sulfates, sulfonates, and succinates. Anionic softeners are used in specialized areas of application where the physiological activity is low, e.g., medical textiles. As the majority of fluorescent brightening agents are anionic, they are widely used in conjunction with anionic softeners for the resin finishing of white cellulosic fibers [3].

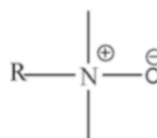
Reactive softeners contain functional groups capable of reacting with particular functional groups in some fibers [8] (e.g., the -OH group of cellulosic fibers). This type of finishing is permanent to washing due to the formation of covalent bonds between the softener and substrate.



Finish, Textile, Fig. 5 N-methylol derivative (R: C₁₇H₃₅)



Finish, Textile, Fig. 6 Compound containing silanol group



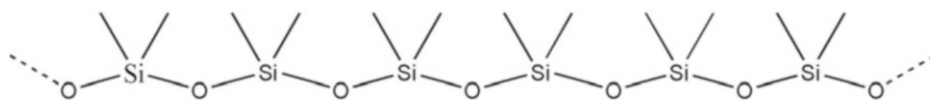
Finish, Textile, Fig. 7 Alkyldimethylamine oxide softener (R: long alkyl chain)

N-methylol derivatives of stearic acid amides and urea-substituted compounds are very successful as reactive softeners (Fig. 5) [3]. One example of a silicon-based reactive softener is given in Fig. 6, which is a modified siloxane and contains functional silanol groups. The silanol group could potentially be replaced by other functional groups such as amines or alcohols.

Amphoteric softeners have limited textile applications but are very popular in personal care products (e.g., shampoo formulations) due to their low toxicity [3]. One example of an amphoteric softener is shown in Fig. 7.

Water and Oil-Repellent Finishes

There is a distinction between “water-repellent” and “waterproof” fabrics. Waterproofing is achieved, for example, by coating the fabric with rubber. Such a treated fabric will not only be impermeable to water but also most notably against air and perspiration; as such, “waterproof” fabrics are uncomfortable to wear. A water-repellent finish, on the other hand, remains permeable to air and is achieved by the application of



Finish, Textile, Fig. 8 Polydimethylsiloxane

hydrophobic chemicals to the fabric [1]. These types of “water-repellent” finishes are normally used for clothing and will be discussed in this chapter. As very few textiles are inherently water repellent but none are oil repellent, so an additional process must be added to display these properties.

If the internal cohesive interactions within a liquid are lower than the adhesive interactions operating between a solid surface and the liquid, then a drop of the liquid will spread on the solid surface. Based on this theory, repellent finishes (both oil and water) work on the principle of reducing the free energy of the fabric surface [8]. In this context, it should be mentioned that alongside the chemical composition of the material, the geometry of the textile surface also plays a significant role in the wettability of the textile [1]. It has long been recognized that superhydrophobic surfaces require a unique combination of two fundamental properties, namely, surface roughness and low surface energy [2]. The functional chemicals used to achieve water-repellent finishing significantly vary in their chemistry, their role being to reduce the surface energy of the fabric by adding lower surface energy chemical groups to the surface [3].

The chemicals traditionally used to achieve water repellence may be divided into a few groups: metal salts, soap/metal salts, wax, pyridinium-based finishes, organometallic complexes, *N*-methylol derivatives, silicone finishes, and fluorochemical finishes [3]. Among these, only fluorocarbon finishes are known to repel both oil and water [3, 12], whereas the other finishes can repel only water. Moreover, perfluorinated derivatives are effective at very low concentrations [1]. Polymeric fluorochemical finishes are typically acrylic or methacrylic polymers with perfluoro side chains.

Silicon-based water repellents are also used (Fig. 8).

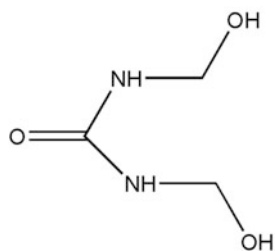
Easy-Care and Durable Press Finishes

As the name suggests, “easy care” and “durable press” are applied to textiles to impart minimum care properties, mostly to cellulosic fibers and their blends with other fibers. Various terms have been used to describe this area, including easy care, easy-to-iron, no iron, crease resistant, wrinkle resistant, wrinkle free etc. However the technically correct term is “cellulosic anti-swelling” or “cellulosic cross-linking” finishes [8].

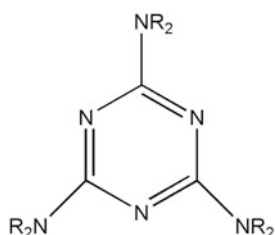
Cellulose (natural or regenerated) is a linear polymer formed by the condensation of β -D-glucopyranose that contains 1 $\rightarrow$ 4-glycosidic linkages. The presence of hydroxyl groups along each chain creates extensive H-bonding both between (intermolecular) and within (intramolecular) the chains. Formation of such intramolecular H-bonds imparts “stiffness” to the cellulose molecules by restricting movement of the 1,4- β -D-glucopyranose units. The -OH groups are also responsible for the hydrophilicity of cellulosic fibers; owing to the high density of the -OH groups, cellulosic fabrics shrink in water and crease upon drying. The main function of easy-care/durable press finishing is to overcome shrinkage and wrinkling by cross-linking the -OH groups.

Although finishing agents are unable to penetrate the crystalline regions of the fiber, the easy-care/durable press finish needs to be of small M_r and preferably high reactivity to enable it to penetrate the amorphous region of the fibers [3].

It was demonstrated in the late 1920s that thermosetting resins (e.g., phenol-formaldehyde, urea-formaldehyde) could impart crease resistance to cellulosic fabrics. Since then, numerous publication and patents have described novel



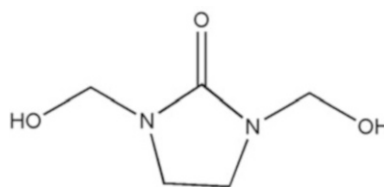
Finish, Textile, Fig. 9 Dimethylolurea



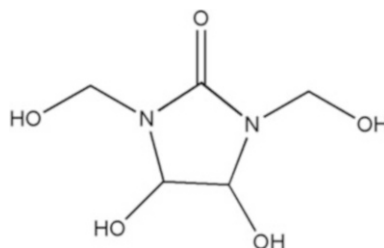
Finish, Textile, Fig. 10 Methoxymethylmelamines (R=CH₂OCH₃, CH₂OH)

cross-linking reagents based on urea-HCHO chemistry [4]. This original type of easy-care products (urea-HCHO resins) used to contain high free formaldehyde, e.g., dimethylolurea (DMU) which is prepared from an excess of HCHO (Fig. 9). These nonreactant types of finishes are only durable to washing temperatures up to 60 °C. Wash fastness can be improved by using reactive resins such as methoxymethylmelamines (Fig. 10).

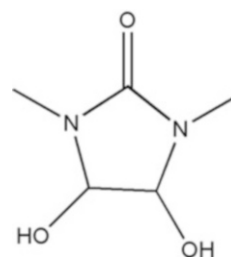
The first cyclic urea-reactant finishing agent was dimethylol ethylene urea (DMEU) which displays poor fastness to chlorine bleaching and adversely affects the light fastness of the finished fabric (Fig. 11). Cyclic urea-based finishing agents, *N,N'*-dimethylol-4,5-dihydroxyethylene urea (DMDHEU), comprise ~90% of the easy-care/durable press finish products that are available in the market [8] (Fig. 12). Due to the presence of two hydroxyl groups at the C-4 and C-5 positions, the reactivity of DMDHEU is low, and therefore, reaction requires a catalyst. Although DMDHEU-containing products contain >0.3% free formaldehyde, the search for formaldehyde-free finishing agents has spanned several decades [3] due to the toxicological effects



Finish, Textile, Fig. 11 Dihydroxymethyl ethylene urea



Finish, Textile, Fig. 12 *N,N'*-dimethylol-4,5-dihydroxyethylene

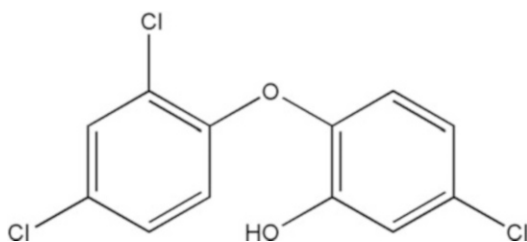


Finish, Textile, Fig. 13 *N,N'*-dimethyl-4,5-dihydroxyethylene urea

associated with formaldehyde. Textiles containing a high level of formaldehyde can give rise to eczema and allergic reactions; furthermore, HCHO is a suspected human carcinogen [8]. Expensive formaldehyde-free finishing agents such as *N,N'*-dimethyl-4,5-dihydroxyethylene urea (DMdDHEU) (Fig. 13) is used for textile processing. DMdDHEU is less reactive than DMDHEU and therefore requires a stronger catalyst and harsher reaction conditions for successful cross-linking with cellulosic fibers.

Antimicrobial Finishes

Antimicrobial finishes are used to inhibit the growth of or to destroy microscopic organisms [3]. This could be for hygiene purposes, i.e. to



Finish, Textile, Fig. 14 2-(2,4-dichlorophenoxy)-5-chlorophenol

protect the user from pathogenic or odor-causing microorganisms or to protect the textile material itself from damage caused by mold, mildew, or rot-producing microorganisms. Formaldehyde is a widely used biocide and preservation product. As discussed in the previous section, the majority of easy-care/durable press finishes contain HCHO, and therefore, these finishes display a small antimicrobial effect. However, for effective antimicrobial action, specific chemical finishes are used. Traditional antimicrobial agents such as copper naphthenate, copper-8-quinolate, and numerous organomercury compounds are strictly regulated because of their toxicity and potential for environmental damage [8].

A common antimicrobial agent extensively used for cellulosic material is polyhexamethylene biguanide (PHMB) which has long been used in applications such as cosmetics and swimming pools. Along with conventional textile applications, PHMB can also be used in the production of medical textiles. PHMB is cationic and therefore displays excellent durability on cellulosic material [3]. Triclosan (Fig. 14) is another antimicrobial agent that is used extensively in mouthwashes, toothpastes, deodorants, and on textiles. This is nontoxic to humans and used as durable antimicrobial finish on polyester and polyamide fibers and their blends with cotton and wool. Heavy metals such as silver, copper, and mercury can provide antimicrobial effect in the form of the metal or metallic salt. Antimicrobial high-performance textile fibers (e.g., polyester, nylon, etc.) can be produced that include nanoparticles of heavy metals (e.g., silver). Chitosan is a natural

biodegradable polymer which is nontoxic and shows microbial resistant.

Conclusions

As a result of increasing demands for superior quality and functionality in textiles, the sophistication of textile finishes and finishing operations has increased. New types of finishing are being developed to provide novel effects such as fragrance finishes, well-being finishes, bionic finishes, as well as finishing for smart textiles, medical textiles etc. Parallel to the search for novel finishes, existing finishes (and application techniques) are constantly being scrutinized to ensure that processing is as efficient as possible from technical, economical, and ecological perspectives. Textile finishing can be expected to remain a highly valuable tool which can significantly enhance the value of the finished textile product.

Cross-References

- [Colorant, Textile](#)
- [Coloration, Textile](#)
- [Dye](#)
- [Dye, Functional](#)

References

1. Fiscus, G., Grunenwald, D.: *Textile Finishing: A Complete Guide*. Editions High Tex, Sausheim (1995)
2. Marsh, J.T.: *An Introduction to Textile Finishing*, 2nd edn. Chapman & Hall, London (1966)
3. Haywood, D. (ed.): *Textile Finishing*. Society of Dyers & Colourists, Bradford (2003)
4. Vigo, T.: *Textile Processing and Properties: Preparation, Dyeing, Finishing, and Performance*. Elsevier, New York (1994)
5. Jackman, D., Dixon, M., Condra, J.: *The Guide to Textiles for Interiors*. Portage & Main Press, Manitoba (2003)
6. Carty, P., Byrne, M.S.: *The Chemical and Mechanical Finishing of Textile Material*, 2nd edn. Newcastle upon Tyne polytechnic products Ltd, Newcastle upon Tyne (1987)

7. Hargrave, H.: *From Fibre to Fabric: The Essential Guide to Quiltmaking Textiles*. C & T Publishing, Lafayette (2008)
8. Schindler, W.D., Hauser, P.J.: *Chemical Finishing of Textiles*. Woodhead Publishing Limited, Cambridge (2004)
9. Tesoro, G.C.: Textile finishes. *J Am Oil Chem Soc.* **45**(5), 351–353 (1968)
10. Daoud, W.A., Xin, J.H., Tao, X.: Superhydrophobic silica nanocomposite coating by a low-temperature process. *J Am Ceram Soc.* **87**(9), 1782 (2004)
11. Balu, B., Breedveld, V., Hess, D.W.: Fabrication of “roll-off” and “sticky” superhydrophobic cellulose surfaces via plasma processing. *Langmuir*. **24**, 4785 (2008)

Fitting (Anglo-Saxon English)

► [Luminaires](#)

Fixture (American English)

► [Luminaires](#)

Flame Light Lamps

► [Combustion Lamp](#)

Focal Colors

► [Color-Opponency, Unique Hues](#)

Forecast

► [Color Trends](#)

Four-Dimensional Color Vision

► [Tetrachromatic Vision](#)

Full Radiator

► [Blackbody and Blackbody Radiation](#)

Functionality of Color

Antal Nemcsics¹, Malvina Arrarte-Grau² and Galen Minah³

¹Budapest, Hungary

²B.Arch (R.Palma U.) M.L.A. (Edinburgh U.)

Environmental Colour Design Study Group (AIC), Lima, Peru

³Department of Architecture, College of Built Environments, University of Washington, Seattle, WA, USA

Synonyms

[Color effects to humans and the environment](#);
[Color roles](#); [Color uses](#)

Definition

The use, informative, and aesthetic impact of color on humans and the environment.

Color and Function

Human environment may be considered as a system, since man and environment elements are mutually determinant in any sense – an interrelation concerned with demand of man for his environment. The function of the built environment is based on social demand, necessity raised to a social level.

Within the system of man and elements of his environment, structural relations are defined and

Antal Nemcsics: deceased.

sustained by complex functions. This complex is composed of three types of functions: utility, aesthetic, and informative ones. Utility function is understood as the designation and purpose of environment elements. Aesthetic functions bring about properties of environment elements which enable one to experience utility functions. Informative functions involve properties of environment elements by which their functions, uses, and operation become understandable for man. These preliminaries underlie the present day approach to environment color design, considering the totality of utility, aesthetic, and informative functions as the system of demands.

Environment color design has to serve the expression of the complex function of environment elements. Various functions of color-bearing environment elements are strictly interrelated and are prone to change into each other, a characteristic to be considered in the methodology of environment color design as a design process. It is necessary to develop design methods suitable for simultaneous and differentiated consideration and coordination of these components [1–3].

Color function components of environment color design as a design process can be deduced from the man to color relations. The man to color relation cannot be regarded in an abstract way. Colors are always associated to some object, phenomenon, or process, implicating into this relation their complex functions. In this respect, the theory of environment color design has been concerned with the possibility to express function by the color of the environment element. It can be stated, for instance, that more saturated colors of the color space, with longer dominant wavelengths, act dynamically and hence suit to express functions involving dynamism. The less the saturation of these colors, the lower their dynamism. The greater the hue, saturation, or lightness differences between members of color complexes to express a function, the more dynamic is the function expressed by them. The intensity of the expression of function is more affected by the variation of lightness differences than by saturation differences. Again, it is found that a function is expressed not so much by a single color but by a complex of several colors. Accordingly, the

expression of function by the environment elements contributes to harmony relations of colors borne by the element, and Coloroid color harmony relations may be combined with function expression indices [3, 4].

Disclosure of the rules of function expression cannot rely solely on visual information, since it is physiologically possible that stimuli perceived by one of the sensory organs should create perceptions normally transmitted by another sensory organ.

Utility Function of Colors

Environment is the space for human activities aimed at satisfying human demands. Since the beginnings of society, architecture has been expected to meet human demands. Analysis of and reckoning with functions necessarily took place in every period, even if functional demands have significantly changed and developed through the ages. The scientific analysis of functions and a conscious, integrated attention to all the requirements could not, however, develop earlier than in contemporary age, at a higher level of social development. This has led from early functionalism, the narrow interpretation of function, to an extended, wide-range functional approach, to the complex consideration of the integer system of relations inclusive of demands, requirements, and the color environment.

Of course, this functional approach itself is undergoing development, although in its germs it has been present in every stage of development throughout the history of architecture and, recently, in environment color design. Nowadays, however, it has become more conscious, unambiguous, and scientific, and it has risen to be an important factor of the up-to-date approach. The relationship between structure and function in architecture has been recognized, and this knowledge unavoidably affected environment color design. The ideas of environment color designers have to be directed by modern scientific thinking.

Activities in various areas of the built environment have become extremely differentiated. Functions have become so manifold that

categorization is not only difficult but may lead to misunderstandings. Nevertheless, functions need to be categorized; otherwise the relations between functions and colors cannot be delineated, and the role of colors in the utility function of the environment cannot be studied.

According to the activities accommodated, the environment can be divided into working spaces, community spaces, living spaces, and traffic spaces.

To demonstrate the usable function of colors in different institutions, hospitals will be dealt with in the next paragraphs. Functions of hospital buildings are very special. For some people, a hospital provides accommodation for a shorter or longer but nevertheless limited time, while for others it is a working place. Therefore, the demands of both groups have to be respected not only concerning construction and equipment but also coloration. The more so, since these demands are sometimes contradictory [2, 5, 6].

In most cases, the state of mind of a patient is *a priori* somewhat abnormal. The lonely, worried patient, so to say forcedly dragged from his habitual environment, is in need of being addressed and protectively accommodated by his new environment. On the other side, the work of the medical staff requires a great deal of concentration, and excessive color effects may distract their attention or even disturb his work. These anomalies have led to the absurd dispute among color dynamists as to whom should priority be given in the hospital ward.

In a hospital, honoring of the utility functions can be done according to different aspects. Coloration should help orientation. Applying a certain color for doors of rooms with identical or similar functions may eliminate much uncertainty. Doors not for patients may be colored so as to convey this information. Proper color design may express connections between rooms. Any obtrusive color sensation has to be avoided in examination rooms and surgeries. In bathrooms, natural color effects of the sun, sand, and water have to be mimicked. The patient finds his picture in the mirror pale and sickly if it is being colored by near strong reflections from a green surface. A similar pale-colored impression arises when the patient observes his

face neighbored by fiery orange or fiery red colors. But in rooms frequently contaminated by blood, green should be applied as complementary to red. Ward walls possibly need cheerful, warming, approaching, but not very saturated, colors. Eventual chromotherapeutic effects have to be exploited. A febrile condition in itself yearns for a cool environment. In a condition of depression, a multicolored environment has to be preferred. In a labor ward or an intensive care unit, a positively reassuring coloration has to be applied.

Aesthetic Function of Colors

Just as any work of art, every element of the environment, and the environment itself, is an inseparable unity of content and form. Just as the human environment is a space for human functions, environment elements serve somehow the realization, accomplishment, and completion of human functions, and by all that, they fulfill their own function. Thus, the human environment is up to its aesthetic function if it expresses its utility function in conformity with unity of content and form – where content is the utility function and form is the shape and color of the elements [1, 3, 5].

The essential condition of an aesthetic effect is unity between content and form, as so defined. Furthermore, the objects forming the content of the environment, or indeed the entire environment, are essentially functional. It follows that the integral expression of content can be realized by the harmonious functioning of all the components of the entire space. The practical-utility and spiritual-mental components of function are interdependent. The mental arises, and is inseparable, from practical. Now it is not realistic any longer to believe that the aesthetic form of an object or space should be possible without knowing its function. Neither is it believed that there exist universal aesthetic formulae applicable in every case nor that it is possible to force upon the environment any fine color harmony without knowledge of its function. In environment color design, it is also necessary to consider how much importance is attributed to its practical functions from the point of view of human life as a whole.

That is to say, any work and activity has its emotional, mental, and spiritual attachments. Therefore each object, tool, or space claims some of these attachments, depending on its role, importance, and function in human life.

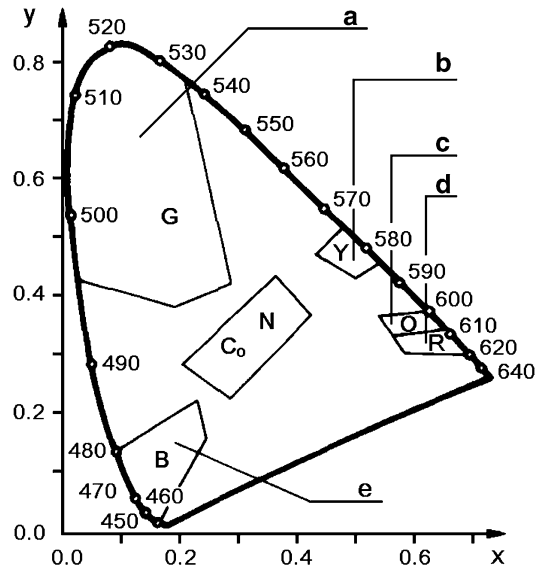
Form becomes a necessity in the course of a “function experience,” imprinting on consciousness a harmony sensation as inseparable unity between content and form. Of course, the sight of a color complex may give aesthetic pleasure, but when separated from the content and function of the space, this pleasure is rather superficial, as it does not offer a full space sensation.

To be able to create a form expressing the message of the built environment, the designer has to be acquainted with relationships between environment structures, the so-called composition relationships, such as that of material-function-form, conditions of space effects, or the role of space-time-motion in an environment.

Informative Function of Colors

Information functions of environment serve to interpret for man the functions, the ways of utilizing, and the operation of the environment and its elements. An important part of the informative functions of the environment is expressed by color information. Depending on the message, color information may be interpreted as logic information or as artistic or aesthetic information.

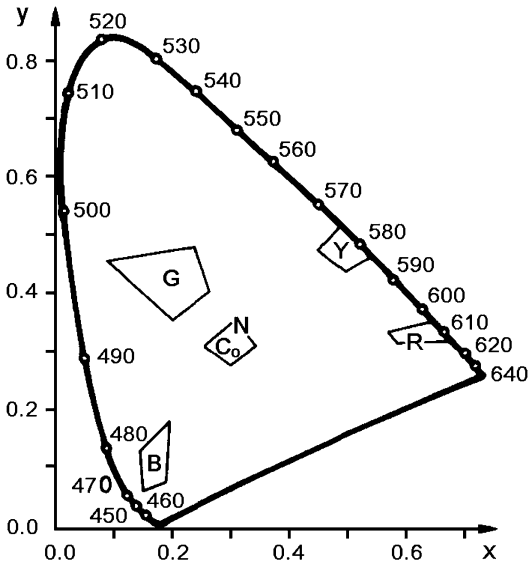
Logic and aesthetic information are carried by the same elements, but a different structure belongs to every form of message. They can be characterized partly by their different visual systems, by the differentiation of their complexity and structure, and partly by the psychical differences of their messages. A message is transmitted by highlighting, contracting, and grouping some visual codes in the information-bearing surface or space while disregarding others. A group of colors attracts attention when it excels by regularity and its structure is well recognizable. Recently, the analysis of these relations has been tackled with the methods of semiotics. Although these studies mostly concerned other than visual structures, the results can also be applied to color dynamics.



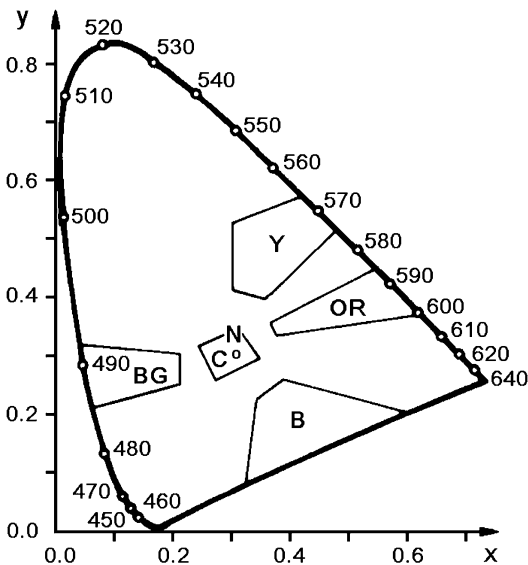
Functionality of Color, Fig. 1 Safety signal colors in the CIE diagram: (a) information, (b) warning, (c) prohibition, (d) fire, (e) instruction. *N* neutral complementary domain, *Y* yellow, *O* orange, *R* red, *B* blue, *G* green

Logic information is conveyed by standard codes; they are practical and strictly mental and transmit knowledge. They prepare decisions of receivers and control their behavior and attitudes. Logic information is transmitted, e.g., by internationally agreed safety color signals. Locations of these colors in the CIE diagram are seen in Fig. 1. For instance, green means information, orange warning, red prohibition, and blue instruction. A special field of the built environment is traffic. Traffic signal colors are plotted in Fig. 2. Recently, colors indicating various technological processes have been standardized, as seen in Fig. 3. A special field of technology is the color signals of pipelines. Color signals for the fluid carried have been standardized, including different uses of the same fluid. For instance, there are different signal colors for drinking water, for utility hot water, condensing water, hot water for heating, soft water, neutralizing sewage, and other applications. In Fig. 4, pipeline colors are plotted in the given Coloroid sections [3, 4].

Color information of aesthetic content is mostly emotional. It expresses inward conditions and the wish to exert mental and emotional effects



Functionality of Color, Fig. 2 Traffic light colors in the CIE diagram: *Y* yellow, *R* red, *B* blue, *G* green, *N* neutral complementary domain



Functionality of Color, Fig. 3 Technology signal colors in the CIE diagram: *Y* yellows, *OR* orange reds, *B* purples and violets, *BG* bluish greens, *N* neutral complementary

based on a common semantic knowledge. Because of their operative and recording functions, visual codes are not only carriers of the message, and social ideas of the color-designed space as a work of art, but may also display the

attitude typical of the artistic subject and culture. It follows from all these that color dynamics creates complex color conditions and puts them into a colored world; thereby it has a manifold message, that is, it both expresses and molds human consciousness and emotions. In other words, environment color design is an artistic activity on a large scale, a piece of art existing in the built space, and as such, it has not only to cope with its specific iconic task but also to meet the special conditions of color dynamics.

Functionality in Architectural Color

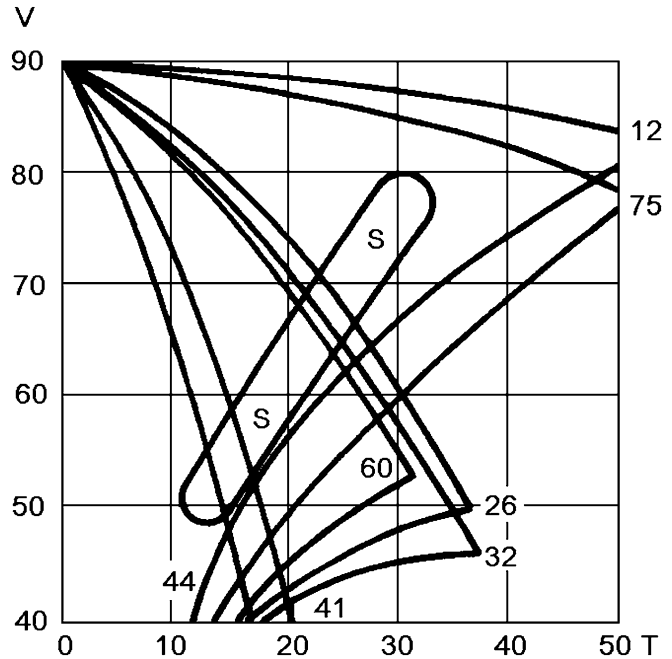
Color is essential to the appearance of architecture and, therefore, impacts on the visual outcome of the built environment. Encompassing the practical and aesthetic aspects of architecture, the colors of building exteriors generate physical and emotional effects on users, and in addition, serve as means of expression. For its breadth of implications, architectural colors, which once were simple to resolve, these days may be considered a matter for experts.

Before taking color decisions for building exteriors, orientation and climate should be observed, especially when critical conditions of visibility and reflectivity occur. Color, as a primary perception, produces physiological effects on humans. Because of the surface extensions involved, architectural colors are relevant to ergonomics, specifically concerning visual comfort. For instance, color may be used to compensate the glare produced by large vertical areas in environments of high luminosity.

The physical properties of color account for energy conservation and thermal comfort in buildings, complementing design principles of orientation, insulation, and ventilation. Following these criteria, light colors and reflective materials are used to diminish the overheating of roofs in hot climates, whereas opaque black surfaces seek to absorb maximum heat. Besides solar radiation, the outermost layer of a building should withstand rain, humidity, dust, and contamination. Durability and maintenance are important considerations for deciding exterior colors, as cladding and

Functionality of Color,

Fig. 4 Signal colors (*S*) to distinguish pipelines in Coloroid sections. The horizontal axis shows the Coloroid saturation scale, the vertical axis, and the Coloroid lightness scale. Numbers represent Coloroid hues



painting materials come in colors which age differently.

Apart from the practical aspect, color on buildings exerts power at different levels of cognition with impact on the user's emotional well-being. Clarity of perception, orientation, understanding of general and specific information, appreciation of beauty and place identity are only some of the advantages that may be experienced through architectural colors in this respect.

From the User's Perspective

Colors may be used on façades and exposed structures in pursuit of specific intentions of owners, architects, and authorities. Image, which refers to the character of a building within its context, is usually prioritized when selecting colors for exteriors. Color is often used as a means of decoration, serving aesthetic purposes but also acting as a reference to ownership, identity, and content.

Architectural colors may allude to a symbolic aspect, as hues convey meaning. Using this as a resource, the chromatic composition of a building may attract attention for personal or commercial

interests. For instance, corporate identity and affiliation may be represented on a façade by replicating colors associated with a certain brand or institution.

Codification applied to façade colors may also contribute to orientation, as distinctive tones and combinations are perceived in the distance. In this way the role of a building is anticipated by its external colors, provided that these are in accordance with its use. As an example, the colors of a day-care center give a clear idea that it is directed towards children.

On a broader scale, familiarization with large areas such as residential developments and commercial centers may be clarified through color design. On the other hand, in areas that conserve features of natural and cultural value, it is important to respect color as an element of identity [7], for it grants uniqueness and sense of permanence to places.

Building exteriors act as a backdrop to human activities, especially in cities where scale and monotony are not welcoming. Colors on façades promote visual interaction with the surroundings and also present an opportunity to enhance outdoor spaces. As a component of the public realm,

these are exposed to acceptance or rejection from the public. By maintaining a balance between consistency and variety, tradition and renovation, and beauty and fun, the aesthetic aspect of architectural color becomes relevant in the development of a positive attitude and a sense of belonging in the user.

From an Architectural Perspective

The spatial and material characteristics of a building provide the support for architectural color. Color and light determine the perception of three-dimensional form, affecting the appreciation of distance, size, and definition of planes in a way that the building and its color produce an integral result, as a sculptural object.

The colors of buildings may be the outcome of constructive processes, as usually happens with exposed construction materials. Nowadays a variety of industrial products for exterior cladding is offered in specific color palettes, or customized, as in the case of paint finishes. Anyhow, to accomplish the various functions of color in architecture, and meet the requirements of each project in the semantic, formal, and contextual aspects, an understanding of color properties and of the relevant cultural codes is necessary.

As regards the building's image, enhancement is a desirable objective. Color may reinforce its style and design concept. In this way, an edifice of architectural or historical value may recover its authenticity by color restoration. Addressing the parts of a building with colors that suit the actual function of the building is a straightforward approach to color design. Otherwise, existing constructions may be transformed to convenience through plain colors and textures, or else, by the application of pictorial techniques such as supergraphics [8], trompe l'oeil, and camouflage. However, the preservation of the architectural character is conditioned to the appropriateness of color design and the particularities of each case.

On a broader context, by the use of color, it is possible to modify the relation of hierarchy, proximity, and perceived area among buildings and its connection to the setting. In urban and natural

environments, color has the potential to regulate the impact of obtrusive structures. More so, when inserting a new building in an existing network, a mindful color proposal is required. Color may be applied to buildings by using principles of attachment and detachment [8], seeking color effects similar to those of the surroundings with the purpose of integrating the constructions to the panorama, or else, to turn them into landmarks, by color differentiation.

Color applied to architecture offers countless possibilities. These should be carefully studied, selected, and refined. For the sake of places, in the best interest of the user and to avoid color pollution, architectural color should be handled in a knowledgeable way.

Color Tectonics in Architecture

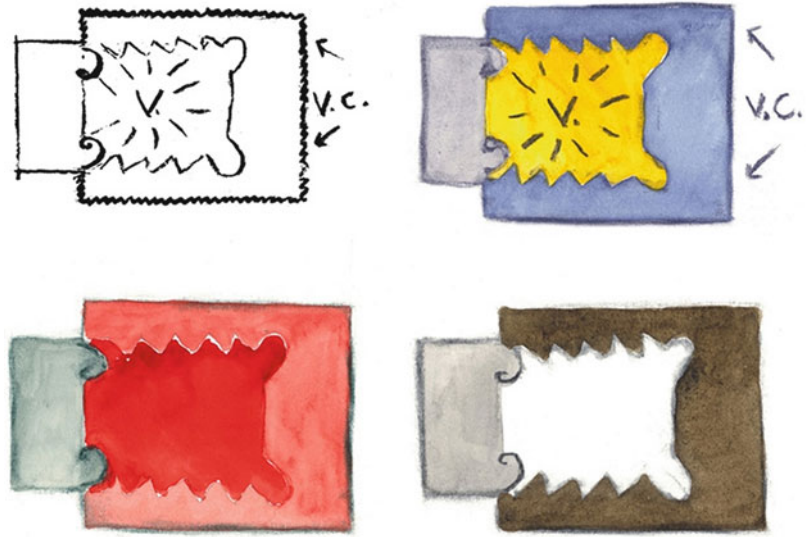
Tectonics in architecture and urban design is the discipline dealing with the principles of design, structure, construction, and ornamentation. Color tectonics refers to the function that color plays in accentuating these principles.

Color and form are codependent in built form. In architecture color plays many roles throughout the design process. Color is used dynamically in diagrams in the conceptual phase. In the design development phase, color can enhance the perception of physical form and define interior space. In detailing color can express the parts of the building that contribute to the whole. These color and form strategies will be present as material color in the finished building. In the architecture, color can be functional as an enhancement to form or abstract as expressive media. Color can be associative by carrying cultural values and meaning, as well as eliciting emotional response. In all cases color combined with form becomes another layer of interpretation and clarification of the intentions of the designer [8].

Conceptual Phase

A building, like a city, has an order based upon a clear hierarchical relationship of parts to the

Functionality of Color,
Fig. 5 Color adding
 nuance to a concept
 diagram by Louis Kahn



F

whole. Architecture begins with a concept expressed both verbally and in diagram. In the conceptual phase of the design process, line drawings are used to represent an abstract relationship of the essential parts of the building. These parts can be described metaphorically or formally, and the relationship of these parts to one another creates the generative idea that is the point of departure for the design. These relationships are often expressed as a dialogue between oppositions such as public/private or active/passive. They can also represent events in the experience of architecture such as hierarchy, separation, connection, transition, and assimilation. The drawings are usually monochromatic, but if one assigns colors to these diagrams representing the character of the part (i.e., red represents active functions and blue passive, or saturated hues might be dominant and muted hues subordinate), then color juxtaposition and contrasts can set up a more visually dynamic relationship of these parts in the diagram. These color choices in the conceptual phase are chosen for their dynamic relationships in juxtaposition rather than by functional criteria, but these colors will often influence choices in the design development phase such as the delineation of three-dimensional form, building materials, and lighting.

A repertory theater is a place where actors must project their voice to the audience without

acoustical support. A design concept for a repertory theater by Louis Kahn was first expressed metaphorically as “a violin within a violin case.” The diagram was a box representing the outer shell of the building or the “violin case” with an irregular form inside representing the “violin,” an acoustically designed theater. In black and white, this is just a simple diagram. With color, however, the violin and violin case form an expressive dialogue between these parts, making them more dynamic and meaningful (Fig. 5).

Design Development

In the design development phase of a design process, color is used to accentuate or diminish three-dimensional form. The exploration of perceptual effects created by color, particularly the spatial effect of color, color and atmosphere, and the familiarity with the principles of camouflage [9] are helpful tools in this pursuit (Fig. 6). The color decisions here are based on how the eye perceives color and form. The accentuating and disguising of form with color are the techniques supporting color tectonics. Using color in models in this phase establishes color juxtapositions and contrasts that will be present in the final building. Model studies similar to the experiments by Lois Swinoff [10] are valuable exercises.

Functionality of Color,

Fig. 6 Transforming two solid cubes and a void with a color illusion

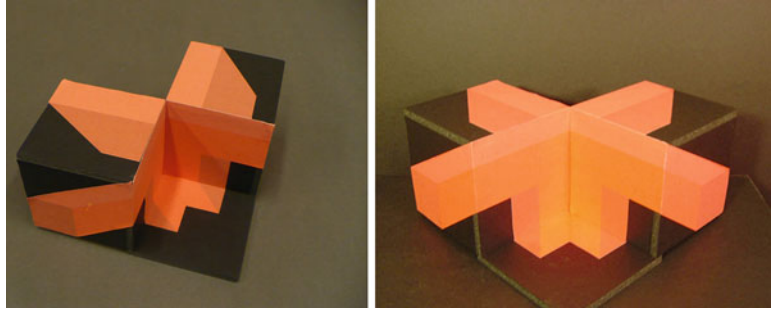
**Functionality of Color,**

Fig. 7 Color used for “deconstruction.” (Portland Building by Michael Graves, Portland, Oregon)



The purpose of color tectonics is to use color to clarify and enhance the perception of building form. This form is made up of parts arranged hierarchically that unite to shape the whole. This color/form carries the aesthetic and expressive intentions of the architect, and clarity of expression is usually the goal. In some instances, however, there is an intention to disguise and actually deconstruct the true form, and color becomes a powerful tool in this endeavor. This was the case

in the Portland Building during the period known as deconstruction in architecture (Fig. 7).

Detail

The details of buildings carry expressive content. Structural details communicate what elements in the building are supporting and load bearing and what are non-load bearing and serve to enclose the



Functionality of Color, Fig. 8 Color contrast emphasizing structural detail. (The Arts United Center by Louis Kahn, Fort Wayne, Indiana)

building and define space within it. Many functional parts of a building are repetitive, and through this repetition patterns and rhythms are established which support the unity of the building as a whole. Color contrasts provide a powerful tool for making these parts visible in a meaningful way. In the main entry facade to the Fort Wayne Repertory Theater, light gray concrete is contrasted with darker brick to both characterize the structural forces within a large wall and to accentuate the main entry (Fig. 8).

Color Imagery

In the final or building phase of the design process, color choices are more specific. To this point, color has been an integral tool in clarifying the conceptual and design intentions of the architect and not as a secondary consideration in the design process. The final color/form imagery will both convey these intentions and also carry

meaning through cultural associations, symbolism, and emotional response. The use of colors from the natural and built environmental contexts can give the architecture a sense of place as well [11]. In this process color is a principal consideration through all phases. The imagery experienced in the architecture will carry these design ideas in form and structure through perceptual color effects, but there will be an additional layer of meaning with the associative uses of color and the aesthetic imagery from experiencing the architecture as a whole in its environmental context.

Cross-References

- ▶ [CIE Chromaticity Diagrams](#), [CIE Purity](#), [CIE Dominant Wavelength](#)
- ▶ [Color Appearance](#)
- ▶ [Color Dynamics](#)
- ▶ [Color Harmony](#)

- [Color Palette](#)
- [Color Pollution](#)
- [Environmental Color Design](#)
- [Unique Hues](#)

References

1. Pickford, R.W.: *Psychology and Visual Aesthetics*. Hutchinson, London (1972)
2. Birren, F.: Science and art, objective and subjective. *Color. Res. Appl.* **10**, 180–186 (1985)
3. Nemcsics, A.: *Colour Dynamics. Environmental Colour Design*. Ellis Horwood, New York (1993)
4. Nemcsics, A.: The coloroid colour system. *Color. Res. Appl.* **5**(2), 113–120 (1980)
5. Frieling, H.: *Farbe in Kultur und Leben*. Battenberg Verlag, Munich (1963)
6. Nemcsics, A.: Das Farbenpräferenz-Indexzahlensystem im Dienste der farblichen Raum-gestaltung. *E'KME Tud. Kozl.* **13**(1/2), 21–261 (1967)
7. Lenclos, J., Lenclos, D.: *Colors of the World, the Geography of Color*. W.W. Norton, New York (2004)
8. Porter, T.: *Architectural Color*. Whitney Library of Design, New York (1982)
9. Minah, G.: Color as idea. *Colour Des. Creat.* **2**(3), 1–9 (2008)
10. Rao, J.: *Introduction to Camouflage and Deception*. Metcalf House, Delhi (1999)
11. Swirnoff, L.: *Dimensional Color*. Birkhauser, Boston (1989)

Fundamental Colors

- [Primary Colors](#)

Galaxy Color Magnitude Diagram

Michael H. Brill
Datacolor, Lawrenceville, NJ, USA

Definition

A scatter graph of galaxies showing the relationship between each galaxy's absolute magnitude and its estimated temperature, or between optical and perceptual proxies for these quantities. The construction of such a diagram is similar to that of the Hertzsprung-Russell diagram, but the interpretation in terms of physical properties is not as precise.

Absolute Magnitude and Temperature Scales

To compute a galaxy's absolute magnitude, it is treated as a point-like object, whereupon its radiation is corrected (by the inverse-square law) to a distance of 10 pc, and the absolute magnitude is computed as the number of $10^{-0.4}$ attenuations of similarly compensated power from a reference to achieve a match.

The temperature is also the same "effective temperature" as is used in the H-R diagram. However, galaxies are visible from much farther away than individual stars, so a galaxy's recessional redshift

exerts an appreciable influence on its effective temperature. Because light from farther galaxies (and greater redshift) requires more time for light to travel to the Earth, the galaxy color-magnitude diagram, if it were not redshift corrected, would depict an earlier universe at the red end of the temperature scale. However, in practice the temperature is redshift corrected to be in the "rest frame" of the galaxy. In this case, the galaxies divide into two clusters: bright and reddish as opposed to dim and bluish. The existence of such clusters places strong constraints on theories of galactic formation.

History

Galaxy color-magnitude diagrams came into wide use starting in 1990, when the Hubble Space Telescope was launched into orbit. The Space Telescope was able to render high-contrast images outside the Earth's atmosphere without encountering the absorption, scattering, and haze that beset Earth-bound telescopes. In digital images conveyed to earth from the Space Telescope, it was possible to see the colors, not only of stars, but of whole galaxies that are so distant as to be too dim to measure on the Earth.

Coordinate Options

Color-magnitude diagrams are not always defined in terms of fundamental physical properties such as

those described above. In one embodiment, the galaxy color-magnitude diagram coordinates are defined by the log outputs of two sensors, typically called *g* and *r*, the sensors having different spectral sensitivities. The abscissa records *r* (which stands in lieu of the absolute magnitude), and the ordinate records *g*–*r* (which relates to the temperature). At the top of the diagram, one finds the red sequence of galaxies (elliptical galaxies), and at the bottom, one finds the blue cloud (spiral galaxies).

Cross-References

- [Color Magnitude Diagrams](#)
- [Hertzprung-Russell Diagram](#)

Gamut Volume

Philipp Urban

Fraunhofer Institute for Computer Graphics
Research IGD, Technische Universität Darmstadt,
Darmstadt, Germany

Synonyms

CIECAM02 (Standards: CIE); CIELAB
(Standards: CIE)

Definition

The gamut volume describes how much of a color space a gamut occupies. Hence, it depends on the color space in which the gamut is represented. In order to be used as a meaningful quantity to compare gamuts, a representation within a highly perceptually uniform color space is recommended. Color spaces typically used to calculate gamut volumes are CIELAB or CIECAM02 [1]. Please note that these spaces are only nearly perceptually uniform. Hence, spaces with improved uniformity [2, 3] are sometimes used as well mathematically; the gamut volume V is defined by the volume integral

$$V = \iiint_G dc_1 dc_2 dc_3,$$

where G is the gamut, and c_i are the color space coordinates, e.g., $(c_1, c_2, c_3) = (L^*, a^*, b^*)$. The gamut volume is expressed in *cubic color space units* and mostly rounded to thousands for device and media gamuts [4].

In a perceptually uniform color space, the gamut volume is nearly monotonically related to the number of distinguishable colors within the gamut, i.e., the larger the gamut volume, the larger is the number of distinguishable colors in the gamut. Mathematically, the number of distinguishable colors n within the gamut may be expressed as the smallest number of just noticeable distance (JND) spheres covering the gamut, i.e.,

$$n = \min \left\{ |\mathbb{S}| \mid G \subset \bigcup_{S \in \mathbb{S}} S \right\},$$

where G is the gamut, $\mathbb{S}$ is a set of JND spheres, and $|\mathbb{S}|$ are the number of elements in $\mathbb{S}$. It is worthy to mention that the JND depends on the viewing condition, i.e., the luminance level, viewing field, adaptation state, etc. Hence, the gamut volume and the number of distinguishable colors within the gamut are not two sides of the same coin. However, assuming similar viewing conditions, gamut volumes can be reasonably compared to determine the gamut that includes more distinguishable colors.

The gamut volume is only one quantity to compare gamuts and does not give any information about colors covered by the gamut. A gamut that includes highly relevant colors for a distinct application should be preferred over a gamut that does not cover these colors regardless of the gamut volume.

Calculation

A major problem of accurately calculating the gamut volume is the non-convex shape of the gamut within a perceptually uniform color space. For output devices, such as printers, displays, or

projectors, a good approximation of the gamut is given implicitly by an accurate device model

$$\mathcal{M} : P \rightarrow C,$$

i.e., a predicting function that maps the control value space P (e.g., $P = \text{RGB}$, CMY) into the color space C used for gamut representation (e.g., $C = \text{CIELAB}$, CIECAM02 , etc.). The gamut G can be parametrized by the device's control values, i.e., $G = \mathcal{M}(P)$. Using this parametrization and assuming that $\mathcal{M}$ is an injective differentiable function with continuous partial derivatives, the gamut volume can be computed using the substitution rule for integration

$$\begin{aligned} V &= \iiint_{G=\mathcal{M}(P)} dc_1 \, dc_2 \, dc_3 \\ &= \iiint_P |\det(J\mathcal{M}(p_1, p_2, p_3))| dp_1 \, dp_2 \, dp_3, \end{aligned}$$

where $J\mathcal{M}$ is the Jacobian matrix of $\mathcal{M}$. Since P has typically cubic shape, the latter integral can easily be calculated numerically. Furthermore, no representation of the rather complex gamut boundaries within C is required.

While mathematically elegant this method is often inapplicable, e.g., for image gamuts or if the control value space of a device has more than three dimensions, such as for multi-ink printers (CMYK, CMYKOG, etc.). In this case, a gamut boundary descriptor (GBD) [5] describing the gamut boundary by a tessellation within C can be utilized. This tessellation is then used as the boundary of a tetrahedralization that fills the entire interior of the gamut. In most well-behaved cases, tetrahedralization can be simply generated by forming tetrahedra with each of the surface triangles and one inner point, e.g., the centroid of the GBD vertices [4]. To approximate the gamut volume, the volumes of all tetrahedrons within the tetrahedralization are added up. The volume of a tetrahedron with vertices a , b , c and d is given by $1/6|\det(-a - b, b - c, c - d)|$. For devices using additive

color mixture, such as displays, the gamut is convex in the CIEXYZ color space. A tetrahedralization of the gamut in CIEXYZ space allows a convenient partitioning into sub-gamuts for further consideration. Since a transformation,

$$\mathcal{F} : \text{CIEXYZ} \rightarrow C,$$

from CIEXYZ into a perceptually uniform color space C can be assumed to be a diffeomorphism, the following equation applies:

$$G = \bigcup_{T \in \mathbb{T}} \mathcal{F}(T),$$

where G is the gamut in C , and $\mathbb{T}$ is the set of all tetrahedrons within the tetrahedralization. Hence, the gamut volume can be computed as follows [6]:

$$\begin{aligned} V &= \iiint_G dc_1 \, dc_2 \, dc_3 \\ &= \sum_{T \in \mathbb{T}} \iiint_{\mathcal{F}(T)} dc_1 \, dc_2 \, dc_3 \\ &= \sum_{T \in \mathbb{T}} \iiint_T |\det(J\mathcal{F}(X, Y, Z))| dXdYdZ, \end{aligned}$$

where $J\mathcal{F}$ denotes the Jacobian matrix of $\mathcal{F}$. It should be noted that the meaningfulness of the computed gamut volume – with respect to its relation to the number of distinguishable colors – is usually more effected by the uniformity of the color space C rather than by the computational method.

Applications

Most often gamut volumes are used to compare imaging systems with each other or with standard gamuts (e.g., sRGB or Adobe RGB). Also the volume of gamut intersections is used and related to standard gamuts, e.g., to state that a display's gamut covers more than 95% of the Adobe RGB gamut.

The gamut volume is sometimes used as an objective function to optimize imaging systems (e.g., displays or printers). Particularly, the spectral or colorimetric characteristics of printing inks or display primaries are optimized to maximize the gamut volume. Most of such optimizations are subjected to other beneficial properties of imaging system or feasibility constraints.

References

1. CIE Publication No. 159: A Colour Appearance Model for Colour Management Systems: CIECAM02. CIE Central Bureau, Vienna (2004)
2. Thomsen, K.: A euclidean color space in high agreement with the cie94 color difference formula. *Color. Res. Appl.* **25**, 64–65 (2000)
3. Urban, P., Schleicher, D., Rosen, M.R., Berns, R.S.: Embedding non-euclidean color spaces into euclidean color spaces with minimal isometric disagreement. *J. Opt. Soc. Am. A* **24**(6), 1516–1528 (2007)
4. Morovic, J.: *Color Gamut Mapping*. Wiley, Hoboken (2008)
5. Morovic, J., Luo, M.R.: Calculating medium and image gamut boundaries for gamut mapping. *Color. Res. Appl.* **25**, 394–401 (2000)
6. Rodriguez-Pardo, C. E., Sharma, G., Speigle, J., Feng, X.-F., Sezan, I.: Efficient computation of display gamut volumes in perceptual spaces. In: *IS&T/SID, 19th Color and Imaging Conference*, pp. 132–138. San Jose (2011)

Ganglion Cells

Arne Valberg¹ and Jan Kremers²

¹Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

²Department of Ophthalmology, University Hospital Erlangen, Erlangen, Germany

Definition

Retinal ganglion cells are the first real neurons in the visual system, and they transmit the visual information to the brain. The retina processes the visual signals, and they determine the response properties of the retinal ganglion cells and the constraints for color vision.

Physiology of Retinal Ganglion Cells

This entry is only concerned with primate retinal ganglion cells. The morphology and physiology of ganglion cells in other mammals and vertebrates may be quite distinct. Trichromatic theories of color mixture have had a tendency to over-emphasize the role that L-, M-, and S-cone receptors have for color *perception*. The realization, however, that these physiological units can only explain the laws of additive color mixture and color matching (metamerism) has led to an end of this bias. Modern cone-opponent color vision models assume antagonistic inputs of two or more cone types to the receptive field of retinal ganglion cells (e.g., +L-center –M-surround, giving an “L–M” ON-center cell, or an +M-center–L-surround, “M–L” ON-center cell; also called increment cell). The opposite configuration is found in “–L + M” and “–M + L” OFF-center cells (decrement cells). Ganglion cells with S-cone inputs can be characterized as “M–(S + M)” and “S–(L + M)” cells [1–5]. This organization is partially combined with spatial antagonisms of the receptive field centers and surrounds. The idea of parallel pathways has moved the scientific interest from the ancient idea of cones as color units toward retinal ganglion cells as the basis of color vision.

Retinal ganglion cells are situated near the inner surface of the retina. They receive cone signals modified by horizontal, bipolar, and amacrine cells. In the ganglion cells slow, standing potentials of the cones and bipolar cells are transformed into a series of nerve impulses (spikes). Spikes are “all or none” electrical signals of about 1 ms duration each. The number of spikes per second is used as code for transmitting information. Excitation of a cell results in an increased firing frequency, whereas inhibition results in a reduced rate of firing. The number of spikes of a particular cell contains information about the intensity or strength of a visual stimulus. Information about a stimulus’ compound features such as contrast, form, color, movement, direction of motion, distance, etc. is conveyed by activation of more or less specialized cell types.

Retinal ganglion cells have long axons extending into different brain areas. The projection to the lateral geniculate nucleus (LGN) and from there to the various visual centers of the brain is the main pathway that is responsible for conscious visual perception. Ganglion cells are spiking neurons of different sizes, connections, and spectral selectivity. There are about 1.5 million different retinal ganglion cells in the human retina, each type forming a mosaic of receptive fields with a close to ideal covering factor (neighboring dendritic fields do not overlap [6]).

Two major types of ganglion cells carry cone opponent information and are the physiological basis of color vision. Midget retinal ganglion cells are connected in an opponent way to L- and M-cones and have rather small dendritic fields. In the central fovea, the receptive center of a single midget ganglion cell is connected to a single cone. They project to the parvocellular layers of the LGN. It is unclear if the receptive surrounds and the centers of peripheral cells are also cone selective or if they receive mixed cone inputs. With mixed inputs, the cone opponency would mainly originate in the single cone input (for foveal cells) or in a biased input (for peripheral cells) in the receptive field center [7].

Cells making contact with S-cones (the so-called bistratified cells [8] projecting to koniocellular LGN cells) do this in two different synaptic layers and seem to have a coinciding (coextensive) center/surround structure of their receptive fields. Parvocellular (PC)- and koniocellular (KC)-ganglion cells (about 80% of all retinal ganglion cells) have a sustained firing pattern.

In addition to the cone-opponent retinal ganglion cells, the transiently responding, spectrally nonselective magnocellular (MC) or parasol ganglion cells comprise about 10% of all retinal ganglion cells. These contrast-sensitive cells have spectral sensitivities matching the psychophysically determined spectral luminosity function $V(\lambda)$ to luminance contrast. They combine L- and M-cones in an additive way ($L + M$) in both center

and surround of their receptive field. MC-cells also form parallel sub-pathways as they are either of the ON-center, OFF-surround, or OFF-center, ON-surround types. Since these cells have $V(\lambda)$ -like spectral sensitivities and large receptive fields, they are excellent candidates as neural substrates for minimally distinct border (MDB) and heterochromatic flicker-photometric luminance [3].

A relatively small number of retinal ganglion cells are internally photosensitive with their own photopigment, melanopsin, and contribute to the circadian rhythm and pupillary light reflex but little to vision [9]. Recent studies indicate that they may have limited influence on vision including color vision [10].

A particular color percept, say orange of medium lightness, may result in activation and/or inhibition of multiple cell types. The perception and scaling of lightness, for example, may be a result of interaction between nerve processes that are elicited by light increments and light decrements in parallel (increment- and decrement channels) [11]. Equal responses of a low-level cell such as a certain polarization of a cone receptor's potential or the number of nerve impulses/s of a retinal ganglion cell may result from exposure to several different stimuli, such as changes in color, intensity, size, orientation, or combinations of these. The combination of visual attributes that makes up a complex visual object can be regarded as a result of integration, correlation, and comparisons of these lower units' activities. Retinal ganglion cells project via the LGN to specific visual centers of the brain, depending on the "trigger features" they encode.

Cross-References

- [Color Processing, Cortical](#)
- [Cone Fundamentals](#)
- [Magno-, Parvo-, and Koniocellular Pathways](#)
- [Melanopsin Retinal Ganglion Cells](#)
- [Minimally Distinct Border](#)
- [Photoreceptors, Color Vision](#)

References

1. Wiesel, T.N., Hubel, D.H.: Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* **29**, 1115–1156 (1966)
2. DeValois, R.L.: Analysis and coding of color in the primate visual system. Cold Spring Harb. Symp. Quant. Biol. **30**, 567–579 (1965)
3. Lee, B.B., Martin, P.R., Valberg, A.: A physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J. Physiol.* **404**, 323–347 (1988)
4. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 242–265 (1984)
5. Valberg, A., Seim, T., Lee, B., Tryti, J.: Reconstruction of equidistant color space from responses of visual neurones of macaques. *J. Opt. Soc. Am.* **A3**, 1726–1734 (1986)
6. Dacey, D.M.: The mosaic of midget ganglion cells in the human fovea. *J. Neurosci.* **13**, 5334–5355 (1993)
7. Wool, L.E., Crook, J.D., Troy, J.B., Packer, O.S., Zaidi, Q., Dacey, D.M.: Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. *J. Neurosci.* **38**, 1520–1540 (2018). <https://doi.org/10.1523/JNEUROSCI.1688-17.2017>
8. Dacey, D.M., Lee, B.B.: The “blue-on” opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature* **367**, 732–735 (1994)
9. Hattar, S., Liao, H.W., Takao, M., Berson, D.M., Yau, K.W.: Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002)
10. Zele, A.J., Feigl, B., Adhikari, P., Maynard, M.L., Cao, D.: Melanopsin photoreception contributes to human visual detection, temporal and colour processing. *Sci. Rep.* **8**, 2842 (2018). <https://www.nature.com/articles/s41598-018-22197-w>
11. Valberg, A., Seim, T.: Neural mechanisms of chromatic and achromatic vision. *Color. Res. Appl.* **33**, 433–443 (2008)

Ganzfeld

Suncica Zdravkovic
Department of Psychology, University of Novi
Sad, Novi Sad, Serbia

Definition

A ganzfeld (a German word for “whole field”) is obtained when an observer is exposed to an absolutely homogeneous stimulation that covers the

entire visual field. Any particular wavelength or intensity can produce the same effect.

History

The ganzfeld was first introduced in 1930 [1], by German Gestalt psychologist Wolfgang Metzger (1899–1979). Metzger was interested in elementary visual phenomena produced by impoverished visual conditions. The ganzfeld contains no luminance border, luminance ramp, or texture, while the light reaching the eyes is equal from all possible directions. Such conditions, called whiteout, can occur naturally during a snowstorm or in an airplane flying through clouds.

Perceptual Experience in the Ganzfeld

The typical visual phenomenology, for a person of normal sight, includes the emergence of “Eigen-grau” [2], a uniform gray fog of indeterminate depth. This percept occurs regardless of the physical intensity and wavelength of the ganzfeld. Any color in a ganzfeld will eventually fade to filmy gray. The experience of a bright ganzfeld differs from that of a completely dark ganzfeld. Total darkness implies a lack of visual objects and is somewhat less disorienting. However, disorientation in a bright ganzfeld is fast and compelling.

From the beginning, the observer is unable to fixate any part of the field and becomes conscious of the vitreous opacities and blood vessels within the eye. Perceived luminance gradually decreases, the whole field slowly decolorizes, turning into a gray fog. This fog sometimes has a barely noticeable washed-out tone complementary to its actual color [3].

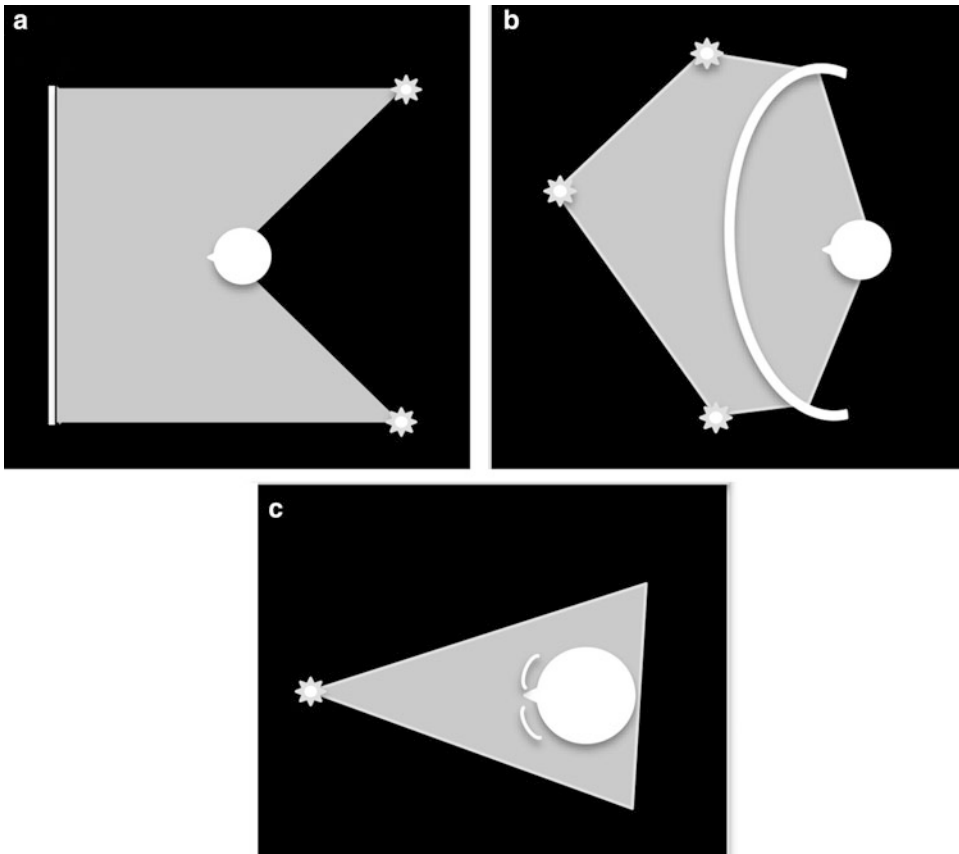
Prolonged exposure to a ganzfeld results in more complex percepts. Occasionally even patterns (dots, lines) are reported. Because hallucinations have been reported in extreme cases, the method has been used outside of vision science. Sometimes prolonged exposure (10–20 min) leads to intermittent losses of vision (so-called “blank-out”) and sometimes participants cannot even report if their eyes are open or closed [3].

In traditional studies, phenomenological reports were obtained from observers trained in providing such introspective reports “post hoc” from memory [3]. The extent of this introspection is debatable, especially in the case of complex visual experience. This constitutes a problem for a phenomenon that might involve disorientation and consciousness-altering states. Alternative methods include “on-demand” reports either elicited by the experimenter or volunteered by the observer. This allows a random sample of momentarily experienced states with no involvement of memory [3]. Unfortunately, some of the same criticisms are applicable here, too. This method also requires training in introspection.

Producing an Experimental Ganzfeld

Although a primitive ganzfeld can be achieved by placing a person in complete darkness, creating a proper ganzfeld of the desired brightness is not a trivial challenge.

In the 1930s, observers were placed in front of either a smooth, homogeneous wall (Fig. 1a) or a large sheet of paper [1, 3]. A simple way to create a ganzfeld is to cover each eye with a section of a table tennis ball (Fig. 1c) [3]. These need to be trimmed to fit the shape of the face surrounding the eye, leaving no extra gaps for illumination to leak in, and they can be attached to the face using cosmetic cement. Industrial production does not guarantee an utterly uniform thickness of the



Ganzfeld, Fig. 1 Traditional spatial arrangements of Ganzfeld [3]. (a) Observer in Metzger’s experiments was placed in front of a uniform wall lit from behind. (b)

Domes lit from outside. (c) Lenses or table tennis balls placed in front of the eyes functioning as light diffusers

plastic, but this is not a problem because the surface is too close to the eye for focus. Another option is placing the head inside a large hemisphere, which is uniformly illuminated from outside (Fig. 1b). A somewhat better solution is the use of translucent contact lenses. This method provides a high degree of homogeneity.

More modern studies tend to use “computer-controlled xenon lamp-based D-ILA projector” creating (usually red) ganzfeld [3, p. 1366].

Finally, any change in luminance or color, including blinking, restores perception. Given that the ganzfeld effect can be achieved fast, this is not necessarily a problem. If prolonged viewing is required, the eyelids can be taped open to prevent blinking, but the eyeballs need to be kept moist. This is easily achieved if the eye is completely covered by the table tennis ball section, as described.

Theoretical Importance

The ganzfeld demonstrates that our senses in general (vision in particular) function to detect change. Invariant stimulation does not produce a sensory response. Framed in Gestalt terminology, an unstructured physical field cannot produce perception due to the lack of perceptual organization.

Observable Characteristics

Brightness. The full ganzfeld phenomenon is achieved on average after 5–7 min [4]. During that time, with no change in stimulation, perceived brightness slowly decreases until it reaches the final value.

If the luminance in a ganzfeld is subjected to subliminal changes over time and space, observers can still detect the correct direction of change [5]. For a completely homogenous ganzfeld (9.1 footlambert), changing 1000-fold per hour at a constant rate, observers detected the direction after 9.7 min.

Wavelength [6]. Although the ganzfeld produces an achromatic, dark gray percept, irrespective of original conditions, the initial

wavelength still dictates some characteristics. Longer wavelengths fade faster than short wavelengths. Shorter wavelengths create an additional sensation of field darkening.

Monocular versus binocular ganzfeld. It has been shown that a ganzfeld presented binocularly (i.e., viewed with two eyes) does not show a loss of visual perception (so-called “blank-out”) typical for monocular viewing [7]. However, recent findings show that a gradual loss of brightness and saturation (so-called “fade-out”) is measurable both for monocular and binocular ganzfeld. Fade-out is however determined by the original field intensity and wavelength [8].

Sex differences [9]. Female observers are more responsive to longer wavelengths. They also preserve visual sensation shorter and report a fewer “blank-out” effects than male observers.

Explanation

The patterns that appear during the short ganzfeld exposures, such as zigzag lines or dot fields, are usually explained by early retinal processes [3]. Receptive and other retinal cells exhibit spontaneous activity, oversaturation, and inhibitory reactions. Prolonged exposure elicits complex percepts, which presumably involves the central nervous system. The loss of vision is probably the consequence of brain discarding uninformative visual input about invariant stimulation.

There is a correlation between “blank-outs” and alpha activity in the resting EEG [10]. Alpha activity (8–12 Hz) is otherwise typical for relaxed states with no stimulation and closed eyes.

Related Phenomena

Aftereffects. If the ganzfeld is induced by yellow table tennis balls, after their removal the world will tend to look bluish, due to chromatic adaptation. Also during the initial bleaching of the original ganzfeld color, sometimes the percept does not remain neutral but adopts a pale complementary color.

Artificial scotomata. Eye movements can be counterfeited by fixating a large portion of the visual field to stimulate an identical part of the retina. This leads to “fade-out” and loss of percept within seconds. To maintain normal perception, a constant change of stimulation is necessary and this is achieved by eye movements. The absence of eye movements in large areas of the visual field, known as artificial scotomata, leads to perceptual filling in, creating large uniform surfaces.

Sensory deprivation. Although the ganzfeld is exceptionally impoverished stimulation, the phenomenon represents perceptual but not sensory deprivation. In the former case, we have simply unstructured stimulation, while in the latter, stimulus’ intensity is extremely reduced.

Dark field vision. When observers are placed in total darkness in a relaxed state, with the eyes closed, they often report the appearance of phosphenes, believed to be the consequence of spontaneous firing of the cells [3]. Phosphenes are similar to the percepts in the early stages of ganzfeld exposures.

References

1. Metzger, W.: Optische Untersuchungen am Ganzfeld. Psychol. Forsch. **13**, 6–29 (1930)
2. Knau, H., Spillmann, L.: Failure of brightness and color constancy under prolonged Ganzfeld stimulation. SPIE Proc. **2657**(4), 19–29 (1996)
3. Wackermann, J., Pütz, P., Allefeld, C.: Ganzfeld-induced hallucinatory experience, its phenomenology and cerebral electrophysiology. Cortex. **44**(10), 1364–1378 (2008)
4. Knau, H.: Thresholds for detecting slowly changing Ganzfeld luminances. J. Opt. Soc. Am. A. **17**, 1382–1387 (2000)
5. Schubert, J., Gilchrist, A.L.: Relative luminance is not derived from absolute luminance. Invest. Ophthalmol. Vis. Sci. **33**(Suppl.), 1258 (1992)
6. Gur, M.: Color and brightness fade-out in the Ganzfeld is wavelength dependent. Vis. Res. **29**(10), 1335–1341 (1989)
7. Bolanowski, S.J., Doty, R.W.: Perceptual blankout of monocular homogeneous fields (Ganzfelder) is prevented with binocular viewing. Vis. Res. **27**, 967–982 (1987)
8. Gur, M.: Perceptual fade-out occurs in the binocularly viewed Ganzfeld. Perception. **20**(5), 645–654 (1991)
9. McGuinness, D., Lewis, I.: Sex differences in visual persistence: experiments on the Ganzfeld and afterimages. Perception. **5**(3), 295–301 (1976)
10. Cohen, W., Cadwallader, T.C.: Cessation of visual experience under prolonged uniform visual stimulation. Am. Psychol. **13**, 410 (1958)

Geniculate Pathways

- [Magno-, Parvo-, and Koniocellular Pathways](#)

Gestalt Grouping Principles (or Laws)

- [Perceptual Grouping and Color](#)

Ghost Image

- [Afterimage](#)

Glare

- [Light Pollution](#)

Glare-Free High Beam – Adaptive Driving Beam (ADB)

- [Automotive Front Lighting System \(Status 2020\)](#)

Glints

- [Texture Measurement, Modeling, and Computer Graphics](#)

Global Illumination

Tobias Ritschel
Department of Computer Graphics, MPI
Informatik, Saarbrücken, Germany

Synonyms

Color bleeding; Indirect illumination;
Interreflection; Light bounces

Definition

Global illumination (GI) is illumination reflected more than once inside a scene.

Introduction

After light was emitted, it has to be reflected before it can arrive at the eye of an observer or a sensor. When this reflection occurs once, it is called local or direct illumination. Illumination which has been reflected more than once is referred to as global or indirect illumination (GI). Reflection of light is sometimes called a bounce. If light is reflected two times from a surface, it is called the one-bounce indirect illumination; if it is reflected three times, it is called two-bounce; etc. As physical reflectance is below one, only a finite number of bounces contribute to the scene illumination.

Example

A simple example of GI is best visible when putting a rectangular achromatic cardboard next to a perpendicular chromatic (e.g., red) piece of cardboard. When directional light is cast perpendicular onto the red cardboard (and consequently no light is cast on the achromatic cardboard), the white cardboard is still illuminated due to the indirect and colored (reddish) light from the chromatic (red) cardboard.

Phenomena

Depending on the reflective properties of the surfaces, GI can create different visual phenomena. For Lambertian surfaces (such as the aforementioned cardboard), light is reflected in all directions equally following a cosine distribution. For specular (mirrorlike) surfaces, light is distributed more in directions similar to the mirror direction (a Dirac distribution). The resulting concentration of light on the receiver side is called a caustic.

GI is subject to similar principles that govern local illumination: if a surface blocks the line of sight between a surface point A and another surface point B, no light is exchanged between A and B, and B is said to lie in A's indirect shadow. In the same way, a mirror can reflect indirect light as well as it can be refracted.

Strictly speaking, multiple reflections or refractions are also GI. However, the term GI is mostly used if the reflection is a true convolution that accounts for multiple directions and not only a single refracted or reflected one. At the same time, lighting from extended (area) light sources, such as the sky, is sometimes referred to as GI as well.

GI tends to be more colorful than local illumination, as light is filtered at every reflection, which is never perfectly white. This contributes to the color mood of a scene but can also be a difficulty, e.g., in white balancing where the illuminant is complicated.

Global illumination is present also in scenes without hard surfaces but participating media. Here, local illumination corresponds to single scattering, and global illumination corresponds to multiple scattering.

Formalization

Conceptually, linking all points where light is reflected results in a light path; the process of illumination along such a path is called light transport. The vertices of the light path can be labeled according to the type of reflection occurring. The light vertex is denoted as L , a diffuse reflection as D , a specular reflection as S , and the eye as E . As an example of this notation, "LDE" describes light that was emitted from the light and bounced diffusely once before arriving at the eye; "LDDE"

denotes one indirect diffuse bounce, “LDSE” the reflection of a diffuse surface, and “LSDE” a caustic.

The rendering equation [1] is a formal way to describe global illumination. It defines the radiance L_o leaving a differential surface element at location $\mathbf{x}$ with normal $n(\mathbf{x})$ in direction ω as

$$L_o(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \int_{\Omega} L_i(\mathbf{x}, \omega_i) f_r(\mathbf{x}, \omega_i, \omega_o) < n(\mathbf{x}) \cdot \omega_i >^+ d\omega_i$$

where L_e is the radiance emitted at location $\mathbf{x}$ in direction ω , Ω is the space of directions in the upper hemisphere above $\mathbf{x}$, L_i is the incoming light arriving at $\mathbf{x}$ from direction ω , f_r is the bi-directional reflectance distribution function (BRDF) which is the ratio outgoing radiance and incoming irradiance, and $< \cdot >^+$ is a dot product that is clamped at zero.

The difficulty of solving this equation is due to the fact that the incoming light and outgoing light appear on both sides of the equation. Solving the equation for a given scene means to find L such that the equation holds.

Computation

For computer graphics, GI is more challenging to compute than local illumination. Including GI in image synthesis leads to more realistic and visually appealing results. The color mood in a scene is to a large extent related to GI. Several approaches exist to compute it.

The most general approach is Monte Carlo ray tracing: For every pixel multiple rays that model geometric optics are intersected with the scene [2]. Whenever a ray hits a scene surface, it is continued in a random direction. After a number of hits, the iteration is stopped, and local as well as global illumination is accumulated according to the combined reflection properties at the hit positions and the local illumination at the last hit. This procedure can simulate all light effects but takes long to compute or results in noisy images.

Radiosity takes a different approach and discretizes the scene surface into finite elements and simulates the energy exchange between them by solving a system of equations [3]. Its basic principle is the form factor that describes the light arriving at element A from element B. Radiosity is best at simulating diffuse indirect illumination but can be extended to specular light transport. It does not scale well to large scenes or highly specular materials.

If only very few paths from the light illuminate a location, it will appear noisy using Monte Carlo ray tracing. As a solution, instead of starting light paths at the eye, they can be started at the light source in a process called light tracing.

When paths are started from both the eye and the light and end points of sub-paths are connected, the procedure is called bi-directional path racing [4].

For notoriously difficult light paths, where only a low percentage of all possible paths contribute a lot of energy to an image, the solution to the rendering equation can be found using Metropolis light transport sampling [5]. The approach generates random light paths, which are changed (mutated) such that more relevant paths are explored in greater detail than less relevant paths.

At third alternative is photon mapping [6]. It is performed in two passes. The first one is similar to light tracing but stores the hit locations into a spatial data structure (the photon map). In the second path, rays are traced from the eye, but every time a surface is hit, the illumination is constructed from the local density of photons inside the photon map. Photon mapping is the best established method to render caustics.

Several approximate approaches to compute GI exist. Instant radiosity [7] approximates the indirect lighting using a high number of virtual point lights. Lightcuts [8] is an approach to compute GI from the high number of virtual point light in sublinear time. GI rendering at interactive frame rates can be achieved using graphics hardware [9].

If GI is due to extended light sources, and not due to multiple bounces, Image-based lighting (IBL) is applicable [10]. Here, the light entering a scene from a direction is stored into every pixel of a (high dynamic range) spherical raster image. To use this image for rendering, it is preprocessed once and can later be used to compute illumination easily.

A simplified version of GI is ambient occlusion (AO) [11]. AO is the special case of illumination by a constantly white, infinitely extended directional area light source. The resulting illumination is similar to the one found on a cloudy day and contains only directionless and soft shadows.

Cross-References

- [Non-Photorealistic Rendering](#)
- [Sampling Problems in Computer Graphics](#)

References

1. Kajiya, K.: The rendering equation. *Comput. Graph. (Proc. ACM SIGGRAPH)*. **20**(4) (1986)
2. Pharr, M., Humphreys, G.: *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, San Francisco (2010)
3. Cohen, M.F., Wallace, J.R.: *Radiosity and Realistic Image Synthesis*. Morgan Kaufmann (1993)
4. Lafortune, E., et al.: Bi-directional path tracing. *Proc. Comput. Graph.* 145–153
5. Veach, E.: *Robust Monte Carlo methods for light transport simulation*. Ph.D. thesis., Stanford University (1997)
6. Jensen, H.W.: *Realistic Image Synthesis Using Photon Mapping*. AK Peters, Natick (2001)
7. Keller, A.: Instant radiosity. *Proc. SIGGRAPH* (1997)
8. Walter, B., et al.: Lightcuts: a scalable approach to illumination. *ACM Trans. Graph. (Proc. SIGGRAPH)*. **24**(3) (2005)
9. Ritschel, T., et al.: The state of the art in interactive global illumination. *Comput. Graph. Forum*. **31**(1), 160 (2012)
10. Debevec, P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. SIGGRAPH* (1998)
11. Zhukov, S., et al.: An ambient light illumination model. *Proc. Rendering Tech.* **45–55** (1997)

Gloss Meter

Bor-Jiunn Wen¹ and Chao-Hua Wen²

¹Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Glossmeter](#)

Definition

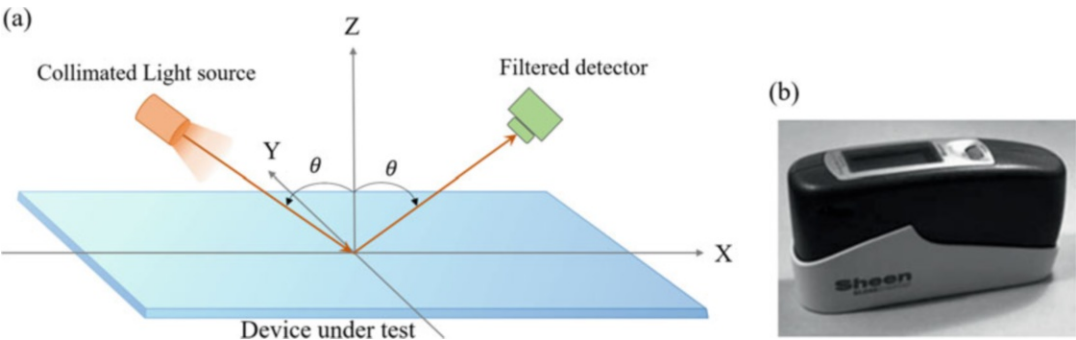
A gloss meter is an instrument that is used to measure specular reflection gloss of a surface. Gloss is determined by projecting a beam of light of intensity at a fixed angle onto a surface and measuring the amount of reflected light at an equal but opposite angle.

Introduction

The gloss meter measures gloss intensity under a fixed set of illumination and viewing conditions [1–3]. Figure 1 illustrates the configuration of the collimated light illumination source and the filtered detector across a limited range of observation reception angles. The measurement results of a gloss meter are related to the amount of reflected light from a black glass standard with a defined refractive index. The ratio between the reflected and incident light for the specimen across a range of angles is compared with the ratio for the gloss standard, is recorded as gloss units (GU).

Applications

There are a number of different geometries available for gloss measurement, according to types of the specimen surface to be measured. For non-metals such as coatings and plastics, the amount



Gloss Meter, Fig. 1 (a) Schematic diagram of gloss measurement and (b) gloss meter

of reflected light increases with a greater angle of illumination, as some of the light penetrates into or is reflected from the surface, which is absorbed or scattered in the material depending on its compositions. Metals have a much higher reflection and are therefore less angularly dependent.

Many international or national standards are available according to different types of gloss and materials such as paint, ceramics, paper, metals, plastics, and so on. For example, ISO 2813 specifies measurements of specular gloss in the paint industry. Many industries use specific gloss meters for their quality control works to ensure consistency in the manufacturing processes. The gloss meter is widely used in the automotive industry, with applications extending from the factory floor to the repair shop.

In the case of transparent materials, the gloss readings can be increased up to 2000 GU due to multiple reflections in the bulk of the material. For these applications, it is common to report the measurement results in percentage reflection against the incident light. For each application, the types of gloss meter and their operation procedures must be strictly specified for intercomparisons. Of the above, the angle of illumination is of most importance. Table 1 defines the measuring angle depending upon three kinds of the different surface characteristics: matte, glossy, and semi-glossy. The gloss range means the initial measurement based on 60° geometry as listed in Table 1. The reading is used to determine which type of geometry to be used. For example, if the initial reading is 40 GU, the result will be 40 GU because

Gloss Meter, Table 1 Gloss types

Gloss type	Gloss range	Measure with
Semigloss	10–70	60° geometry
High gloss	>70	20° geometry
Low gloss	<10	85° geometry

it is within the semigloss range, only using 60° geometry. If the initial reading is 78 GU, a new reading should be obtained using a 20° geometry gloss meter.

Gloss Meter Industry Standards

The following lists various national or international standards for measuring gloss [4]. Many industries adopt the 20°/60°/85° geometries as recommended in ISO 2813 and ASTM D523.

General Gloss Measurement

Standards	Description
ASTM D523 1999 (USA)	Test method for specular gloss. The principal ASTM specular gloss standard. Very similar to ISO 2813
ASTM D3928 1998 (USA)	Test method for evaluation of gloss or sheen uniformity
ASTM D4039 1999 (USA)	Test method for reflection haze of high-gloss surfaces
ASTM D4449 1999 (USA)	Test method for visual evaluation of gloss differences between surfaces of similar appearance

(continued)

Standards	Description
ASTM D5767 1999 (USA)	Test methods for instrumental measurement of distinctness of image gloss of coating surfaces
ASTM E340 1997 (USA)	Test methods for measurement of gloss of high-gloss surfaces by goniophotometry
MFT 30-064 (South Africa)	Local version of ASTM D523
JIS Z8741 1997 (Japan)	Method of measurement for specular glossiness

Paint

Standards	Description
ISO 2813 1994 (International)	Paints and varnishes – determination of specular gloss of nonmetallic paint films at 20°, 60°, and 85°. The principal ISO specular gloss standard. Very similar to ASTM D523
BS 3900:Part D5 1995 (UK)	Methods of test for paints – optical tests on paint films – measurement of specular gloss of nonmetallic paint films at 20°, 60°, and 85°
DIN 67530 1982 (Germany)	Reflectometer as a means for assessing the specular gloss of smooth painted and plastic surfaces
NFT 30-064 1999 (France)	Paints – measurement of specular gloss at 20°, 60°, and 85°
AS 1580 MTD 602.2 1996 (Australia)	Paints and related materials, methods of test – introduction and list of methods
JIS Z8741 1997 (Japan)	Specular glossiness – method of measurement
SS 18 41 84 1982 (Sweden)	Paints and varnishes – measurement of specular gloss of non-metallic paint films at 20°, 60°, and 85°

Plastics

Standards	Description
BS 2782: Pt 5, Method 520A 1992	Methods of testing plastics – optical and color properties, weathering – determination of specular gloss. Similar to ISO 2813

(continued)

Standards	Description
ASTM D2457 1990	Test method for specular gloss of plastic films and solid plastics specifies the primary standard as a perfect mirror with a defined gloss value of 1000. 20°, 60°, and 45°; the 45° method is as ASTM C346 for ceramics

Metals

Standards	Description
BS6161: Part 12 1987	Methods of testing for anodic oxidation coatings on aluminum and its alloys – measurement of specular reflectance and specular gloss at angles of 20°, 45°, 60°, or 85° Reference standard BS 3900: Part D5 (1980); technically equivalent to ISO 7668; replaces BS 1615:1972. At 45°, the dimensions of source image and receptor aperture are as for 60°. Squares with sides equal to the shorter sides of the rectangles are also recommended. Alternatively, total reflection in a 45° prism is used as a reference; source image and receptor aperture are then circular, both with angular diameter $3.44^\circ \pm 0.23^\circ$ (1.5 ± 0.1 mm at a focal length of 25.4 mm)
ISO 7668 1986	Anodized aluminum and aluminum alloys – measurement of specular reflectance and specular gloss at angles of 20°, 45°, 60°, or 85°
ISO 5190	Anodizing of aluminum and its alloys – evaluation of uniformity of appearance of architectural anodic finishes – determination of diffuse reflectance and specular gloss. ECCA T2 (European Coil Coating Association). Specular gloss at 60°

Paper

Standards	Description
DIN 54502 1992	Testing of paper and board; reflectometer as means of assessing the gloss of paper and board
ASTM D1223 1998	Test method for specular gloss of paper and paperboard at 75°. Has

(continued)

Standards	Description
	unusual converging beam geometry. Specifies the primary standard as black glass of refractive index 1.540, not 1.567, at the sodium D-line having a defined gloss value of 100
ASTM D1834 1995	Test method for 20° specular gloss of waxed paper. Another unusual converging beam geometry, different to the previous one
TAPPI T480 OM-90 1990 (USA)	Specular gloss of paper and paperboard at 75°. Same text as ASTM D 1223
TAPPI 653 1990	Specular gloss of waxed paper and paperboard at 20°. Same contents as ASTM D 1834
JIS Z8142 1993 (Japan)	Testing method for 75° specular gloss

Furniture

Standards	Description
BS 3962: Part 1 1980	Methods of testing the finishes of wooden furniture – assessment of low angle glare by measurement of specular gloss at 85°. Similar to ISO 2813: 1978

Floor Polish

Standards	Description
ASTM D1455 1987	Test method for 60° specular gloss of emulsion floor polish. Reference standard ASTM D 523

Ceramics

Standards	Description
ASTM C346 1987	Test method for 45° specular gloss of ceramic materials. Reference standard ASTM D 523
ASTM C584 1981	Test method for 60° specular gloss of glazed ceramic white wares and related products. Reference standard ASTM D 523 {Sheen}

Fabrics

Standards	Description
BS 3424: Method 31: Part 28 1993	Testing coated fabrics – determination of specular gloss

Cross-References

- [Illuminance Meter](#)
- [Reflectance Standards](#)
- [Standard Measurement Geometries](#)

References

1. Hunter, R.S.: Methods of determining gloss. NBS research paper RP 958. J. Res. NBS. **18**(77), 281 (1937)
2. Jones, L.A.: The gloss characteristics of photographic papers. J. Opt. Soc. Am. **6**, 140 (1922)
3. CIE 015:2018 Colorimetry, 4th Edition, Commission Internationale de L'eclairage, Vienna, Austria (2018)
4. Gloss Meter Industry Standards. Available at: <http://www.gloss-meters.com/GlossIntro5.html>. Accessed 24 Dec 2019

Glossing

- [Translation of Color Language](#)

Glossmeter

- [Gloss Meter](#)

Glow Discharge Lamps

- [Neon Lamp](#)

Goethe, Johann Wolfgang von

Rolf G. Kuehni
Charlotte, NC, USA



J. K. Stieler, 1828

Biography

Goethe was born on August 28, 1749, in Frankfurt, Germany, to a lawyer and the daughter of the mayor of Frankfurt. He studied law in Leipzig and Strasbourg (France). Based on his early fame as a poet and novelist, Goethe, at age 26, was invited by the 18-year-old Duke of Sachsen-Weimar to join his court as an advisor. He moved to Weimar where he remained until his death on March 22, 1832. At times he worked as a diplomat and a public servant for the duke. He became known as the preeminent German poet, novelist, and playwright. The philosophical nature of much of his writing and his battle with Kant influenced several later philosophers, such as Hegel, Schopenhauer, Nietzsche, and

Wittgenstein. Goethe's interests were very broad, including pictorial art and certain aspects of science, most importantly in the latter the metamorphosis of plants, where he displayed a pre-Darwinian point of view, as well as the source and nature of color experience. He considered his *Farbenlehre* (Theory of Colours) [1] the most important of his works.

Goethe's views on science were strongly influenced by the seemingly very complex relationship between objectivity and subjectivity. He believed strongly in experimentation, with variability in conditions and multiple replications, which might lead to objective truths. Circa 1793, Goethe wrote an essay *Der Versuch als Vermittler von Objekt und Subjekt* (The Experiment as Mediator Between Subject and Object) which he shared with his friend, the poet Friedrich Schiller, in 1798 in which he described the value of scientific experiments as a bridge between subjectivity and objectivity and the possibility of establishing via multiple experiments of this kind the existence of objective laws of nature [2]. It was only published in print in 1823.

Major Accomplishments/Contributions

His interest in color began early in his life and received input during his extended journey to Italy in 1787 where he began studying Leonardo da Vinci's manuscripts on painting. His first writings on the subject of color date from 1791 *Beiträge zur Chromatik* [Contributions to Chromatics], 1792 *Von den Farbigen Schatten* (Concerning Colored Shadows), and 1794 *Versuch, die Elemente der Farbenlehre zu entdecken* (Attempt at Discovering the Elements of Color Theory). *Zur Farbenlehre*, in three volumes, was published in 1810, the same year P. O. Runge (well acquainted with Goethe) published his *Farben-Kugel* (Color Sphere). The first volume of *Farbenlehre* presents his description and interpretation of physiological colors, physical colors, and chemical colors; a chapter on the possibly singular natural source of color experience; a chapter about the relationship of color science to

other intellectual domains such as philosophy, mathematics, general physics, natural history, etc.; and one on the general esthetics of color. volume 2 is an extended attack on and critique of the color work of Isaac Newton, its main component being the “unnatural” methodology used by Newton when splitting daylight into its spectral components, replaced by Goethe’s method of investigating the effect of the prism when viewing contrasting boundaries in daylight. The result in the former case is the classical spectrum of colors and in the latter the series of so-called boundary or edge colors (for an informative discussion of boundary colors, see Ref. [3]). Recent complex optical experiments have shown that there is no discrepancy between Newton’s and Goethe’s findings (e.g., Ref. [4]). The third volume consists of an extended, if somewhat prejudiced, review of the history of color science (*Geschichte der Farbenlehre*), beginning with the ancient Greeks and ending with contemporaries of Goethe.

Goethe’s fundamental view of perceived colors was that they are the result of interaction between lightness and darkness, an idea already introduced by Aristotle. Goethe saw it confirmed by his experience with boundary colors when viewing edges of white and black fields through a prism, for example, viewing through it the hexagonal charts of Fig. 1, illuminated by daylight.

Goethe devised a color circle consisting of six hues in three complementary pairs, with a purplish red on top, followed on the left, warm side by orange and yellow, and on the cold right side by violet and blue, with green on the bottom. Colors diametrically opposite are approximately complementary. Such opposing colors were viewed by Goethe as harmonic pairs. He also used their relationships in the circle to demonstrate presumed psychological effects.

The response to Goethe’s *Farbenlehre* was mixed. However, in a commentary of 1892 Hermann von Helmholtz said: “And I for one do not know how anyone, regardless of what his views about colors are, can deny that the theory [Goethe’s] in itself is fully consequent, that its assumptions, once granted, explain the facts treated completely and indeed simply” [5].



Goethe, Johann Wolfgang von, Fig. 1 Partial collection of color-related equipment and materials owned by Goethe in the Goethe-Nationalmuseum in Weimar

The English painter Charles Eastlake published in 1840 a translation of Part I as “Theory of Colours” [6]. It became an important source of information for painters, such as J. M. W. Turner. In an introduction to a modern version of the translation in 1970 (remaining in print today), D. B. Judd wrote: “This book can lead the reader through a demonstration course not only in subjectively produced colors (after images, light and dark adaptation, irradiation, colored shadows, and pressure phosphenes), but also in physical phenomena detectable qualitatively by observation of color (absorption, scattering, refraction, diffraction, polarization, and interference)” [7].

References

1. von Goethe, J. W.: Zur Farbenlehre, 3 vols. Cotta, Tübingen (1810)
2. von Goethe, J.W.: Gedenkausgabe der Werke, Briefe und Gespräche, Bd. 16, pp. 844–855. Artemis Verlag, Zürich (1949)

3. Bouma, P.J.: *Physical Aspects of Colours*, 2nd edn. St. Martin's Press, New York (1971)
4. Rang, M.: Goethe's *Farbenlehre* und ihre technische "Aufrüstung" – nicht gegen Newton, sondern mit Newtonscher Optik. In: Wolf Schmidt, G. (ed.) *Colours in Culture and Science*, pp. 246–273. Tredition, Hamburg (2011)
5. von Helmholtz, H.: *Goethes Vorahnungen kommender naturwissenschaftlicher Ideen*. Paetel, Berlin (1892)
6. von Goethe, J. W.: *Goethe's Theory of Colours* (trans.: Eastlake, C. L.). Murray, London (1840)
7. von Goethe, J. W.: *Goethe's Theory of Colours* (trans.: Eastlake, C. L.). MIT Press, Cambridge, MA (1970). Introduction by D. B. Judd

Goniochromatism

► Iridescence (Goniochromism)

Goniometer

► Goniophotometer

Goniophotometer

Cheng-Hsien Chen

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Goniometer

Definition

A goniophotometer is an instrument for measuring light distribution in terms of the luminous intensity or illuminance distribution of a light source, a luminaire, a self-luminous display, or a surface by changing the position of a photometer or a goniometer at a clearly defined geometry.

A typical goniophotometer usually consists of a mechanical device or robotic jib to support the luminaire to be measured and a photometer or a spectroradiometer, and allows their relative position to be changed to form a spherical locus [1].

By analyzing the optical data of the definite angle of spectral irradiance, the total luminous flux, spectral radiant flux, and color quantities can be obtained. Spectral radiant flux refers to the sum of all the radiant energy emitted per unit of time, by a light source over a specific range of a spectrum. Spectral radiant flux is the basis for deriving important luminaire parameters such as total luminous flux, radiant flux, luminous efficacy, chromaticity, correlated color temperature, color rendering index, etc.

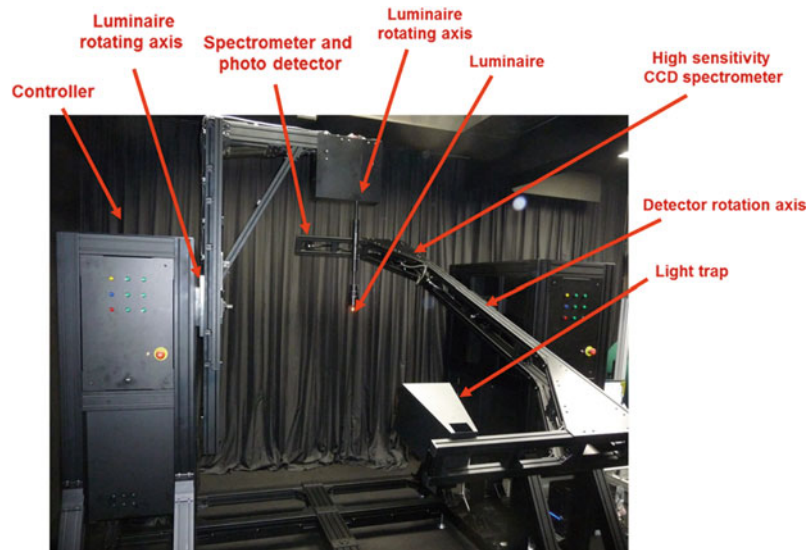
Overview

A typical goniophotometer includes a luminaire and a photometer. There are three common types of goniophotometer, according to the different geometrical arrangements of the two components. The first type of goniophotometer rotates the luminaire along the two assigned perpendicular axes with the photometer fixed. The second type of goniophotometer rotates the photometer instead of the luminaire. The third type of goniophotometer rotates the luminaire along one axis, and its photometer moves along the other axis. The relative orientation of the luminaire and the photometer is then adjusted [2–4].

However, it is complicated to design a goniophotometer, compared with an integrating sphere to measure the luminous flux of the luminaire, because goniophotometric measurements require a large, dark room, a long period for measurement, and great precision of movements. Moreover, for some designs, the luminaire must be rotated, which can cause errors because of its sensitivity to the burning position and draft of the radiant flux. Some luminaire-mounting mechanisms of the goniophotometer can also generate shadow, which affects the luminous flux measured [5, 6] (Fig. 1).

Goniophotometer,

Fig. 1 The International Commission on Illumination type C, spectral-based goniophotometer



The spectral radiant flux measurement goniophotometer uses a spectroradiometer to measure the spectral power distribution of the luminaire at a specific angle. In practice, the spectroradiometer should be calibrated, and the stray light during measurement should be eliminated to achieve measurement accuracy. Stray light is one of the most common sources of measurement inaccuracies for the spectroradiometer. Traditionally, a tunable laser was used to correct the stray light [7]. A novel spectral stray light correction technique for spectroradiometer was then developed. It removes, primarily through an optical frequency comb technique, all stray light that does not fall within a certain fixed wavelength range. This prevents stray light, from indeterminate light sources within the spectroradiometer, from interfering with the measurement process, thus enhancing the accuracy of the measurement [8].

Figure 2 shows a spectral-based goniophotometer, including a detector head module and photodiode array type spectroradiometer set in the jib of the goniophotometer. The luminaire is located in the center of the goniophotometer. The detector head module is arranged according to different predetermined geometries. The key to achieving great

measurement accuracy is to avoid rapid movement, which can cause air flow and vibration.

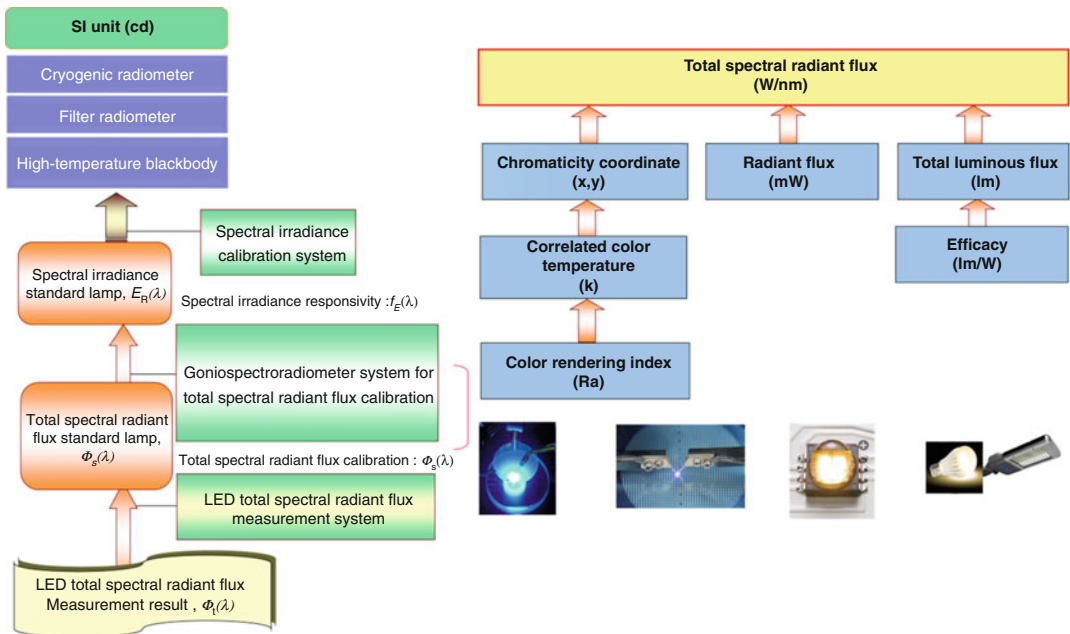
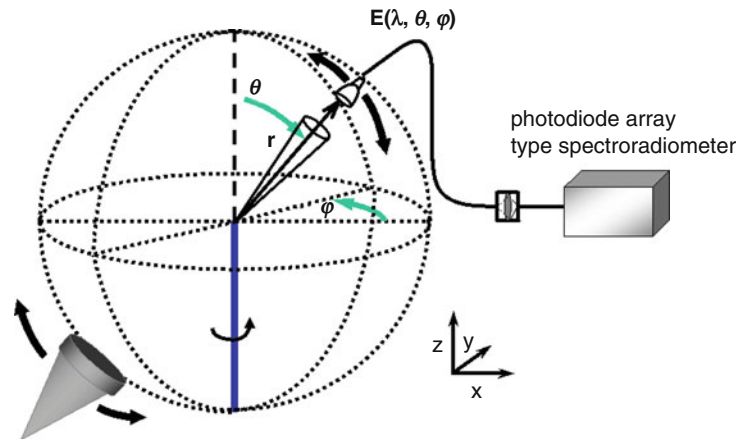
The spectral radiant flux, by integrating the spectral irradiance in the definite angle, may be calculated using the equation below:

$$\Phi(\lambda) = r^2 \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} E(\theta, \phi, \lambda) \sin \theta d\theta d\phi \quad (1)$$

where $\Phi(\lambda)$ is the spectral radiant flux in W/nm unit; $E(\lambda)$ is the spectral irradiance in W/m² nm unit, and r is the radius between the reference plane of the detector and the luminaire.

Light-emitting diode (LED) lighting parameters consist of lighting comfort, product safety, lifespan, and characteristics. Spectral radiant flux is the basis for radiometry and chromaticity parameters and specifications, from which total luminous flux, radiant flux, luminous efficacy, chromaticity, correlated color temperature, color rendering index, etc., can be derived, as shown in Fig. 3. LED total spectral radiant flux is measured using an LED total spectral radiant flux measurement system. The total spectral radiant flux standard lamp carries the total spectral radiant flux and transfers the standard from the goniophotometry measurement system. By using a spectral irradiance standard lamp, the spectral radiant flux

Goniophotometer, Fig. 2 Schematic diagram of a spectral-based goniophotometer



Goniophotometer, Fig. 3 The metrological traceability chain of the total spectral radiant flux standard lamp calibration system

standard is transferred to the irradiance standard and traced to SI units. Therefore, the spectral radiant flux measurement standard can be traced to the SI unit, the candela (cd).

Compared with the sphere-based instruments, a spectral-based goniophotometer (goniospectroradiometer) provides absolute measurement with greater accuracy, than a calibrated flux standard

lamp with a known total spectral radiance. Goniophotometers also report the spatial light distribution data and measure the luminous flux of the particular light distribution of a luminaire, such as a parabolic aluminized reflector lamp (PAR lamp) or a curved lamp. Also, a photodiode array type spectroradiometer should have spectral response calibration, stray light correction,

sensitivity, and linearity certification to ensure accurate measurement.

Cross-References

- [Spectral Power Distribution](#)
- [Spectroradiometer](#)
- [Standard Lamp](#)

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. International Commission on Illumination.: The Photometry and Goniophotometry of Luminaire. CIE 121: 1996, Vienna (1996)
3. International Commission on Illumination.: The Measurement of Absolute Luminous Intensity distribution. CIE 70:1987, Vienna (1987)
4. Illuminating Engineering Society of North America.: Electrical and Photometric Measurements of Solid-State Lighting Products. IES LM-79-08, America (2008)
5. Chen, C.H., Pong, B.J., Jiaan, Y.D., Lin, H.L.: Lamp orientation dependence of an integrating sphere response for directional light source in luminous flux measurement. NCSLI Measure J. Meas. Sci. **8**(2), 42–47 (2013)
6. Lindemann, M., Maass, R.: Photometry and colorimetry of reference LEDs by using a compact goniophotometer. MAPAN. **24**(3), 143–152 (2009)
7. Zong, Y., Brown, S.W.: Simple spectral stray light correction method for array spectroradiometers. Appl. Optics. **45**(6), 1111–1119 (2006)
8. Lui, T.A., Wu, G.N., Hsu, S.W., Pong, J.L., Zong, T.Y.: Optical spectrometer calibration technique based on mode-locked laser comb. Meas. Inf. **154**, 45–48 (2013)

Graininess

- [Texture Measurement, Modeling, and Computer Graphics](#)

“Green-Blindness”

- [Deuteranopia](#)

Guild, John

Robert W. G. Hunt

Department of Colour Science, University of Leeds, Leeds, UK



John Guild was a British scientist who worked at the National Physical Laboratory (NPL) at Teddington in England. Guild spent the majority of his professional life making outstanding contributions to the development of a wide variety of optical instruments and techniques [1].

In the 1920s he wrote several papers, published in the Transactions of the Optical Society, describing the fundamentals of colorimetry which have formed the basis of the discipline ever since.

He measured, for seven observers, the way in which the colors of the spectrum are matched by beams of red, green, and blue light added together. This work, together with a similar study carried out by David W. Wright, with ten additional observers, formed the basis of the international standard for measuring color, the CIE 1931 Standard Colorimetric Observer. The quality of this experimental work was so high that this standard, although now more than 80 years old, is still in universal use. Guild was largely responsible for making this work the basis of the international

system of colorimetry that is still in use today. In collaboration with T. Smith, also of the NPL, he devised the transformation of the experimental results into the XYZ system that is the form in which the standard is used.

In spite of having played a key role in establishing the basis and original standard of colorimetry, after 1931 Guild transferred his interests to other spheres of work and took little further part in the subject. However, he retained his interest in colorimetry and was instrumental in the formation of the Colour Group in the United Kingdom where he was the second chairman of the group (1943–1945) and one of the first honorary members in 1966 [1].

Some of his more important and relevant publications pertaining to this field are listed:

- J. Guild: An equipment for visual spectrophotometry. *Trans. Opt. Soc.* **26**, 74–94 (1924–1925)
- J. Guild: The transformation of trichromatic mixture data: algebraic methods. *Trans. Opt. Soc.* **26**, 95–108 (1924–1925)
- J. Guild: The geometrical solution of colour mixture problems. *Trans. Opt. Soc.* **26**, 139–174 (1924)
- J. Guild: A trichromatic colorimeter suitable for standardisation work. *Trans. Opt. Soc.* **27**, 106–129 (1925–1926)
- J. Guild: A criticism of the monochromatic-plus-white method of colorimetry. *Trans. Opt. Soc.* **27**, 130–138 (1925)
- J. Guild: On a new method of colorimetry. *Trans. Opt. Soc.* **27**, 139–160 (1925–1926)
- T. Smith, J. Guild: The CIE colorimetric standards and their use. *Trans. Opt. Soc.* **33**, 73–134 (1931–1932)
- J. Guild: The colorimetric properties of the spectrum. *Phil. Trans. Roy. Soc. (London) A* **230**, 149–187 (1931)

References

1. The Colour Group (Great Britain): The Advancement of Colour, A History of the First Fifty Years of the Colour Group (Great Britain). The Colour Group (Great Britain), London (1941–1991)

H

Habituation

- [Color Contrast Adaptation](#)

Hair Dyes

- [Coloration, Hair](#)

Halftone Layer Superposition Equations

- [Demichel Equations](#)

Halogen Lamp

Wout van Bommel
van Bommel Lighting Consultant,
Nuenen, The Netherlands

Synonyms

[Iodine lamps](#); [Quartz halogen lamps](#)

Definition

Lamps that produce light as a result of electrical current through a metal wire, contained in a small transparent bulb that heats the metal to incandescence. The bulb also contains halogen that reacts with the vaporized metal, so that bulb blackening by vaporized metal is minimized.

Halogen Lamps

During the operation of an incandescent lamp, tungsten evaporates from the filament and settles on the coldest place inside the lamp (the bulb wall), causing lamp blackening, which leads to a considerable depreciation of the light output during lamp life. In order to keep lamp blackening, and the corresponding light loss, within acceptable limits, bulbs of regular incandescent lamps are relatively large. But in order to be able to operate small-bulb high-light-output incandescent lamps, special measures have to be taken to prevent bulb blackening, which in these small bulbs would quickly lead to unacceptable light losses. The solution to this problem is to introduce halogen into the bulbs. These lamps are called halogen incandescent or, in short, halogen lamps [1, 2]. Thanks to their compactness they are extremely suitable for use in small reflectors to create well-defined light beams. They are widely used for accent lighting and in car headlamps.

However, because of the relatively low efficacy and short lifetime of halogen lamps, the more efficient and longer-life gas discharge and solid-state lamp alternatives have become increasingly more important in all these segments.

Halogen Cycle Principle

One way to eliminate bulb blackening is to add a small quantity of halogen (bromide or iodine) to the fill gas. Under the influence of the hot filament, evaporated tungsten particles chemically combine with the halogen particles. If the temperature is sufficiently high (bulb temperature at least 250°C), the resulting mixture will remain floating in a gaseous state, preventing it from condensing on the coldest part of the lamp bulb to create bulb blackening. The extremely high temperature in the vicinity of the filament causes a chemical reaction that splits the mixture into its original components, tungsten and halogen particles, and the former return to the filament. This phenomenon is called the halogen cycle. It is illustrated in Fig. 1.

A tungsten particle that escapes from one spot of the filament does not return to precisely the same spot. As a result, the lamp will eventually burn out, because there will always be some part of the filament that will become weak over the

course of time, but clearly later than in a regular incandescent lamp. This means that the filament can be heated up to a higher temperature (up to 3000 K instead of 2750 K as in the case of a regular incandescent lamp) while having a longer life. Halogen lamps offer two to five times the life of regular incandescent lamps, rising from 1000 to 2000–5000 h. The higher filament temperature also increases both the luminous efficacy by 10–50% relative to regular incandescent lamps and the color temperature up to 3000 K.

Materials and Construction

Bulb

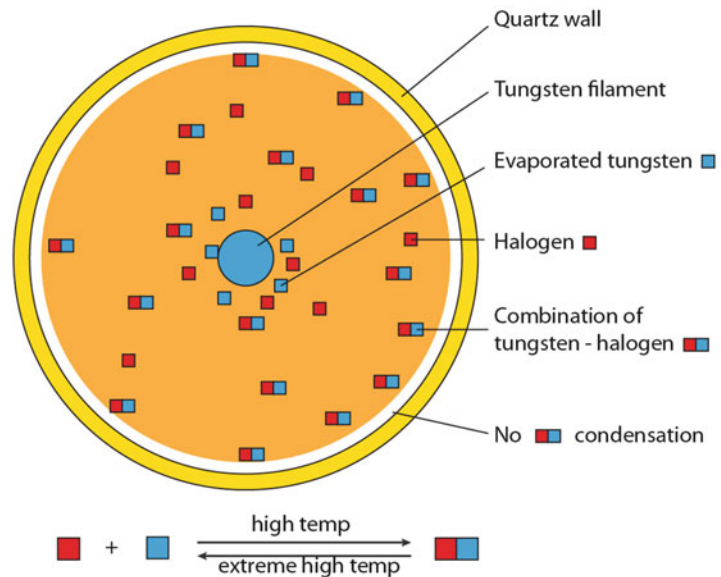
Since a halogen bulb is so small, it becomes so hot that ordinary glass, as used with incandescent lamps, cannot be used because it would melt. Halogen bulbs are therefore made out of quartz, which can withstand high temperatures and mechanical stress. In the manufacturing process, it can be handled similarly to glass (Fig. 2).

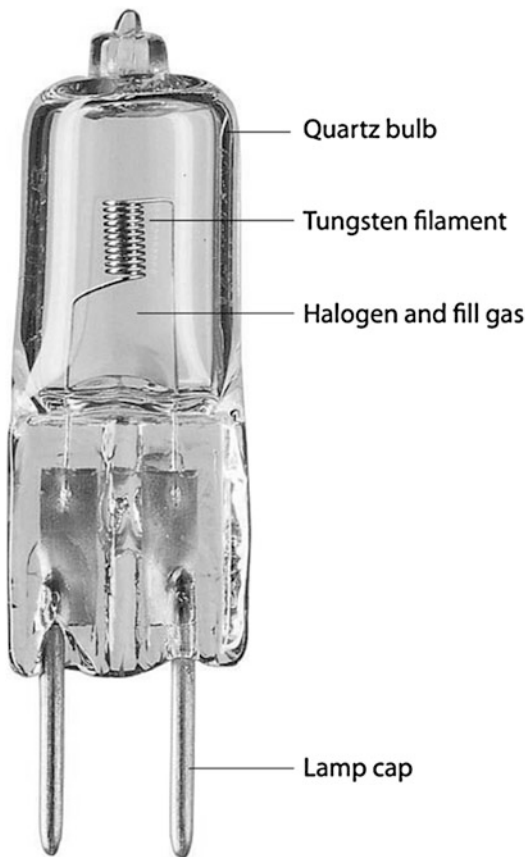
Filament

The single- or double-coiled tungsten filament can be placed axially or transversely in the halogen capsule. The placement has consequences for

Halogen Lamp,

Fig. 1 The halogen cycle
[3]





Halogen Lamp, Fig. 2 Halogen lamp with its components. The bulb can be as small as approximately 10×15 mm

both the efficacy and the light distribution of the lamp.

Fill Gas

As in regular incandescent lamps, a gas filling of krypton or xenon is used to reduce filament evaporation.

Lamp Cap

Halogen lamps are available with a large variety of lamp caps and corresponding bases: two-pin caps and twist caps that ensure that the optical center of the lamp is always in the correct position, ceramic lamp caps for high-voltage, high-lumen-output lamps that become very hot, and “normal” Edison and bayonet caps for high-voltage halogen lamps that can be used just as regular incandescent lamps.

Properties

Energy Balance

Thanks to the higher working temperature of the halogen incandescent lamp, it is more efficient than a regular incandescent lamp. Some 12% of the input power of a halogen lamp is radiated as visible light. Compare this with the 8% of a regular incandescent lamp. The remaining power is lost as heat (through conduction, convection, and infrared radiation).

Luminous Efficacy

At 15–25 lm/W, the luminous efficacy of a halogen lamp is a factor 2–2.5 higher than that of an incandescent lamp. With the infrared-reflecting coating technology (see next section “[Product Range](#)”), the efficacy increases to some 35 lm/W. As with regular incandescent lamps, the lower the wattage, the lower the efficacy. There are halogen lamps where priority has been given to a long life at the expense of a somewhat lower efficacy (e.g., 5000 h and 15 lm/W), and conversely, there are those where the emphasis is on efficacy and not so much on lifetime (e.g., 2000 h and 25 lm/W).

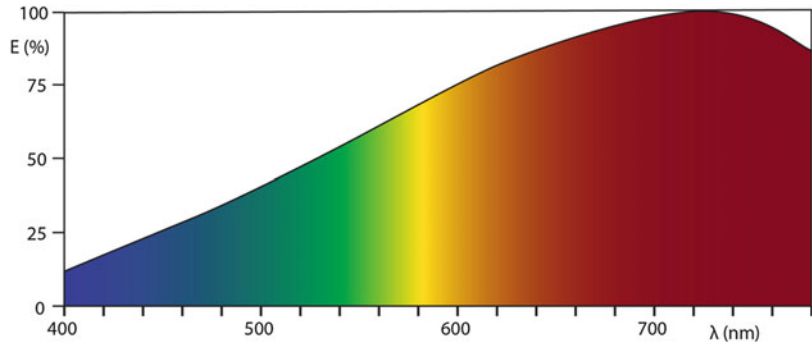
Although the efficacy of halogen lamps is clearly higher than that of regular incandescent lamps, it is relatively low compared with that of gas discharge and solid-state light lamps.

Lumen-Package Range

Common types of halogen lamps are available in the range from some 50 to 2000 lm (corresponding wattage range approximately 5–100 W). Double-ended mains-voltage halogen lamps are available in versions up to some 25,000 lm (1000 W version).

Color Characteristics

Halogen lamps have a continuous spectrum radiating more energy at long wavelengths than at short wavelengths (Fig. 3). Because of the somewhat higher temperature of the filament in halogen lamps compared to that of the filament in regular incandescent lamps, the spectral energy distribution shifts slightly toward shorter wavelengths. Depending on the version (lower

Halogen Lamp,**Fig. 3** Relative spectral energy distribution of a halogen lamp [3]

efficacy/longer life or higher efficacy/shorter life), halogen lamps have a color temperature of between 2800 and 3000 K, respectively. Their color temperature is always slightly higher than that of regular incandescent lamps, thus giving them a somewhat cooler white light. The color rendering index is 100.

Lamp Life

Regular halogen lamps have a lifetime that is at least twice as long as that of regular incandescent lamps, thus at least 2000 h. As explained above, with some halogen lamps, priority is given to efficacy at the expense of lifetime, viz., the 2000 h versions. Versions, where the priority is given to lifetime, can have a life of up to 5000 h.

Lamp-Lumen Depreciation

Thanks to the halogen regenerative cycle, lamp blackening is minimal. Consequently, lumen depreciation with halogen lamps is very small.

Burning Position

With certain exceptions, halogen lamps have a universal burning position. The exceptions are the high-voltage, high-wattage (750 W or more) double-ended types. Here the coiled tungsten filament is so long that with a position that is not near horizontal, the coil would sag so much that the individual coil windings would touch each other, leading to a short circuit.

Run-Up and Reignition

Like regular incandescent lamps, halogen lamps give their full light output immediately after switch on and after reignition.

Switching

The switching behavior of halogen lamps is the same as that of regular incandescent lamps.

Dimming

Halogen lamps can be dimmed in the same way as regular incandescent lamps. Just as with regular incandescent lamps, the phase-cutting system is the more efficient system and is usually employed. Below a certain dimming point, the lamps cool down so much that the halogen cycle will no longer function. From this point on, the filament starts evaporating, just as with a normal, dimmed incandescent lamp. The smaller bulb size of the halogen lamp leads to more blackening in this situation than is the case with regular incandescent lamps. However, since the filament in the dimmed situation is cooler than in the undimmed situation, dimming does not have a real negative effect on lamp life.

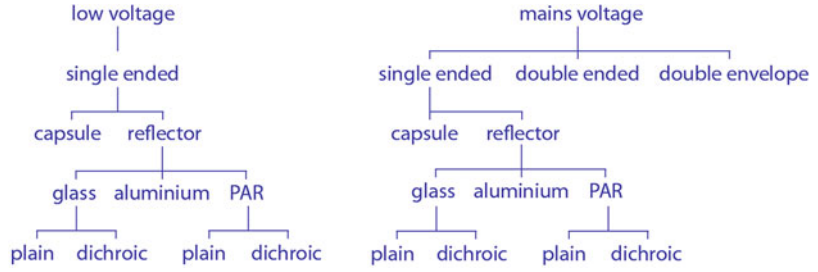
Mains-Voltage Variations

The behavior of halogen lamps as a result of overvoltage is the same as with regular incandescent lamps. Only a few percent of overvoltage results in a drastically reduced lamp life. For example, a permanent overvoltage of 5% reduces lamp life by 50%.

UV Component

Regular incandescent lamps radiate little UV. This is because the spectrum quickly falls off at the short-wavelength part of the spectrum. What little UV that remains is reduced to near zero because regular glass (the bulb material of incandescent lamps) is an excellent absorber of UV radiation. Halogen lamps radiate more UV because of the

Halogen Lamp,
Fig. 4 Range of halogen lamps



Halogen Lamp,
Fig. 5 Single-ended halogen reflector lamps. From left to right: aluminum reflector with antiglare screen, glass mirror reflector, and cool-beam dichroic glass reflector; under pressed-glass, high-voltage, halogen lamp (PAR)



H

higher operating temperature of the filament and because of the fact that regular quartz bulbs, unlike glass bulbs, do not absorb UV radiation. In order to limit harmful or damaging UV radiation from halogen lamps, the quartz used today for most halogen lamps is doped with UV-absorbing material (UV-blocking quartz).

Product Range

Grouping of Halogen Lamps

As far as the operating voltage is concerned, halogen lamps can be placed into two groups (Fig. 4).

- Lamps operating on low voltage (6, 12, and 24 V, where the 12 V versions are most common)
- Lamps operating on mains voltage

Low-voltage halogen lamps are usually single-ended. Mains-voltage versions can be either single-ended or double-ended, according to whether the electrical connection is at one end of the lamp (Fig. 5) or on two separated ends of the lamp (double), as illustrated in Fig. 6.

Single-ended halogen lamps are available both as “naked” bulbs (capsules such as in Fig. 7) and as reflector lamps. The reflector versions exist in three different types: mirror-coated glass reflectors, aluminum reflectors, and pressed-glass PAR reflectors (Fig. 5). The mirror-coated glass reflectors are produced with plain mirror-coating versions and with a dichroic coating (so-called cool-beam type; see later in this chapter).

Some mains-voltage halogen bulbs (both single-ended and double-ended) are housed in an extra outer glass bulb or glass tube with the



Halogen Lamp, Fig. 6 Double-ended mains-voltage halogen lamp



Halogen Lamp, Fig. 7 Low-voltage halogen capsules

familiar Edison or bayonet lamp cap. These are called double-envelope halogen lamps (Fig. 8). They are the more-energy-efficient replacements for regular incandescent lamps.

Cool-Beam Halogen Lamps

A special version of halogen lamps, the so-called cool-beam lamps, makes use of the interference principle. They consist of a small halogen bulb (capsule) housed in a reflector with an interference metal coating, also known as a $\frac{1}{4} \lambda$, or dichroic coating (Fig. 9). This coating splits the radiation coming out of the halogen capsule into an infrared part (heat), which is transmitted through the coating and not reflected by it, and a visible part that is reflected but not transmitted. The practical result is that the mirror coating reflects practically all of the light while allowing the infrared to pass through it to the back of the reflector, taking



Halogen Lamp, Fig. 8 Double-envelope mains-voltage halogen lamps

Halogen Lamp,

Fig. 9 Dichroic interference coating (shown in yellow) in a cool-beam halogen lamp [3]

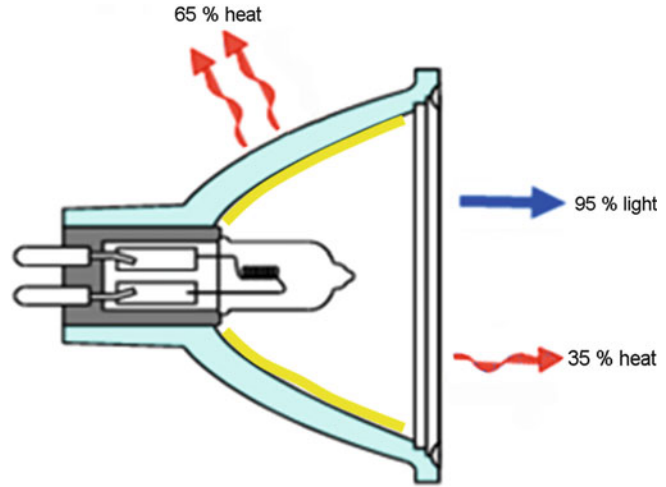
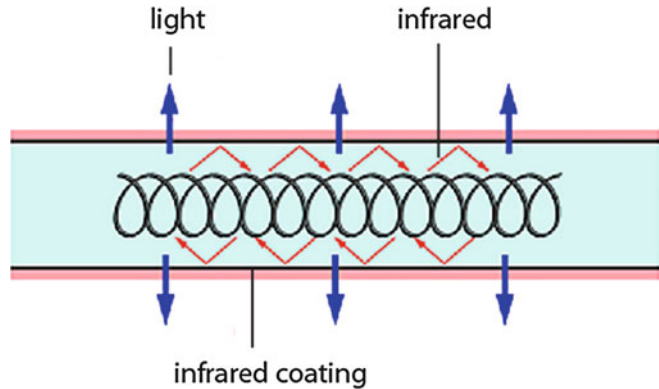
**Halogen Lamp,**

Fig. 10 Infrared-reflective coating based on the interference principle used to reflect the infrared radiation of the filament back to the filament while allowing visible light to pass through [3]



about two-thirds of the heat from the light beam. Thus, only one-third of the generated heat is contained in the light beam. This is why these lamps are referred to as “cool-beam” or “dichroic” halogen lamps. When using this type of lamp in an installation, it is essential to ensure that the heat radiated from the rear of the lamp can be dissipated backwards. The thickness of the coating layer is slightly different for radiation arriving at the coating from different directions. This also means that the wavelength of the radiation where it is split into reflected and transmitted components is slightly different for different directions.

Halogen Lamps with Infrared-Reflective Coating

Another special version of the halogen lamp also makes use of the interference coating principle but

for a completely different reason. Here the reason is thermal recovery: the interference coating applied to the halogen bulb is now so dimensioned that it reflects the infrared radiated from the filament back to the filament while allowing visible light to pass through (Fig. 10). In this way less external energy is needed to keep the filament at the required temperature, thus improving the efficacy of the lamp (by up to 25–35 lm/W). Moreover, the heat contained in the light beam is reduced by approximately 40%.

Cross-References

- [Incandescence](#)
- [Incandescent Lamp](#)

References

1. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
2. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)
3. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)

Hard-Edge Painting

► Color Field Painting

HDR

► High Dynamic Range Imaging

HDRI

► High Dynamic Range Imaging

Helmholtz, Hermann Ludwig von

Rolf G. Kuehni
Charlotte, NC, USA



Image by Ludwig Knaus, 1881

Biography

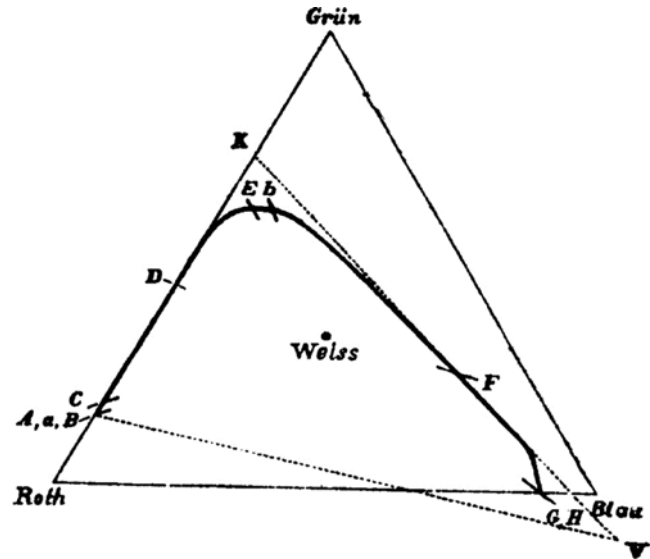
Helmholtz was born on August 31, 1821 in Potsdam near Berlin, Germany, into a well-educated family. His father, an educator, taught him the classical languages and introduced him to philosophy. Helmholtz studied medicine in Berlin under physiologist Johannes Müller, at the same time attending lectures in physics and mathematics. Müller was an adherent of the then broadly accepted philosophical nativism. Helmholtz spent much of his life finding objective arguments against nativism and in favor of empiricism. After spending several years in military service in Potsdam and a stint as associate professor of physiology at the Prussian University, he became a full professor of anatomy and physiology in Bonn. Three years later, he moved to the University of Heidelberg and in 1871 became professor of physics at the University of Berlin, where he remained until his retirement. Helmholtz had a broad range of interests, from astronomy to physics, physiology, and the relationship between the physical world and human sensory perception. In 1847 he wrote an important article on the conservation of physical force [1]. In 1850 Helmholtz invented the ophthalmoscope for inspecting the interior of eyes for medical purposes and revolutionized the practice of ophthalmology [2]. Among Helmholtz's many students were H. R. Hertz, M. Planck, A. A. Michelson, W. Wundt, and W. Kohlrausch (the codiscoverer of the Helmholtz-Kohlrausch effect).

Major Accomplishments/Contributions

In sensory perception, he specifically covered hearing and, most extensively, vision. His most important publication is *Handbuch der physiologischen Optik* (Treatise on Physiological Optics), a 900+-page work whose first edition was published in 1867 when he was in Heidelberg [3]. Its main sections are (1) anatomical description of the eye; (2) physiological optics; (3) dioptrics of the eye; (4) sensations of vision, including simple colors and compound colors; (5) intensity and

**Helmholtz, Hermann
Ludwig von,**

Fig. 1 Chromaticity diagram based on König and Dieterici's fundamental color sensitivity functions determined in Helmholtz's laboratory. Loci of spectral color stimuli are on the curved solid line. Letters indicate loci of Fraunhofer lines in the spectrum



H

duration of the sensation of light; (6) contrast (including colored shadows); and (7) duplicity theory. A second edition authored by Helmholtz was published posthumously in 1896, supervised by Helmholtz' assistant Arthur König who also added over 7800 literature references [4]. A third edition was published in 1909, consisting of the second edition, with contributions by A. Gullstrand, J. von Kries, and W. Nagel. It was translated into English by J. P. C Southhall and published with an additional contribution by the Optical Society of America [5].

Helmholtz built his trichromatic theory of color vision on Newton's findings, Young's theory of three sensor types in the eye, Maxwell's experimental findings of color mixture, and his own extensive experimental data concerning various aspects of color vision. He developed a spectral light mixture apparatus (see portrait) with which, after initial unsatisfactory results causing the mathematician H. Grassmann to challenge his data, he established the first detailed quantitative data for the relationship between color stimuli and percepts. Mathematical models of his findings and concepts were a standard procedure in his work. Under his guidance, his assistants König and Dieterici developed in 1885 a first quantitative version of a trichromatic chromaticity diagram (Fig. 1) [4].

While Helmholtz developed an experimentally supported general psychophysical theory of vision, he was also concerned about and experimented with many additional phenomena of the human visual system. He is the first to mention color constancy, saying that "we always start out forming a judgment about the colors of bodies, eliminating the difference of illumination by which the body is revealed to us" [6]. The color vision theory of Helmholtz, based on Young's conjecture of three sensor types, is known as the Young-Helmholtz theory of trichromatic color vision.

His rival Ewald Hering developed a theory of color vision essentially based on psychology to which he gave a presumed physiological basis. Both views attracted followers [7]. Today, the controversy has been narrowed but still remains. Interestingly, their expressed epistemological views are not as might be expected: Helmholtz viewed lights and colors as symbols. In 1852 he said: "Light and color percepts are only symbols for the relations of reality; with the latter they have as little and as much similarity or relationship as the name of a person or its written form with the person itself" [8]. Hering, on the other hand, believed that, in perception, our access to real objects is a direct one.

References

1. von Helmholtz, H.: Über die Erhaltung der Kraft. Reimer, Berlin (1847)
2. von Helmholtz, H.: Beschreibung eines Augenspiegels zur Untersuchung der Netzhaut im lebenden Auge. Förstner, Berlin (1851)
3. von Helmholtz, H.: Handbuch der physiologischen Optik. Voss, Leipzig (1867)
4. Von Helmholtz, H.: Handbuch der physiologischen Optik, 2nd edn., edited and with contributions of A. P. König. Voss, Hamburg (1896)
5. von Helmholtz, H.: Handbuch der physiologischen Optik, 3 edn., with contributions by A. Gullstrand, J. von Kries, W. Nagel. Voss, Hamburg (1909). English translation with an additional contribution by C. Ladd-Franklin published by the Optical Society of America as Helmholtz's treatise on physiological optics in 1924
6. von Helmholtz, H.: Handbuch der physiologischen Optik, 3 edn., with contributions by A. Gullstrand, J. von Kries, W. Nagel. Voss, Hamburg (1909). English translation with an additional contribution by C. Ladd-Franklin published by the Optical Society of America as Helmholtz's treatise on physiological optics in 1924. English translation, part 2, p. 286/7
7. Turner, R.S.: In the Eye's Mind: Vision and the Helmholtz-Hering Controversy. Princeton University Press, Princeton (1994)
8. von Helmholtz, H.: Über die Natur der menschlichen Sinnesempfindungen. In: Wissenschaftliche Abhandlungen, vol. 2, Barth, Leipzig (1883)

Helmholtz–Kohlrausch Effect

Hirohisa Yaguchi
Professor Emeritus, Chiba University, Chiba,
Japan

Synonyms

Additivity failure for brightness; Brightness-Luminance ratio: B/L

Definition

When a test stimulus of complex radiant power distribution is matched in brightness against a fixed “white” reference stimulus, it is generally found that the luminance of the chromatic test

stimulus is lower than the luminance of the reference stimulus. In other words, a chromatic stimulus of the same luminance as the “white” reference stimulus will, in general, appear brighter than the reference stimulus [1].

This phenomenon is caused by the failure of the additivity law of brightness. Helmholtz first noticed that the additivity law of brightness did not hold, and Kohlrausch studied this in detail, resulting in it being called the Helmholtz-Kohlrausch effect [2].

Overview

Even with chromatic stimuli of the same luminance, colorful stimuli appear brighter than white stimuli. This Helmholtz-Kohlrausch effect is caused by the difference in luminance and brightness. The spectral luminous efficiency function used to define luminance is primarily based on measurements such as flicker photometry, which do not involve chromatic components. On the other hand, brightness is determined by a direct comparison method in which chromatic components affect the brightness criterion, that is, the brightness sensation includes chromatic components in addition to the luminance component. Therefore, even color stimuli with the same luminance vary in brightness appearance, depending on the amount of chromaticness. In addition, the contribution of the chromatic component to the brightness causes an additivity failure in brightness.

Definition of Luminance

The luminance at a point on a surface and in a given direction is the quotient of the luminous intensity in the given direction of an infinitesimal element of the surface containing the point under consideration, by the orthogonally projected area of the surface element on a plane perpendicular to the given direction. According to this geometric definition, luminance is a photometric unit that expresses the perceived brightness of an object.

As a photometric unit, luminance is defined by the following equation with the standard spectral

luminous efficiency function defined by the CIE in 1924:

$$L = K_m \int L_{e,\lambda} V(\lambda) d\lambda, \quad (1)$$

where L means luminance, K_m is the maximum luminous efficacy, $L_{e,\lambda}$ is the spectral radiance, and $V(\lambda)$ is the luminous efficiency function.

Definition of Brightness

Brightness is the attribute of a visual impression according to which a given visual stimulus is presented appears to emit more or less light.

Problems in Photometry

Luminance, which is a unit of photometry, may not always correspond to brightness, which is a visual impression. This problem may be due to two factors in the definition of luminance. First, the luminous efficiency function $V(\lambda)$ itself may be an error. It had been long realized that the weight of the short wavelength region was too low. In 1951, Judd published his modification [3] of the CIE luminous efficiency function now called CIE 1988 $V_M(\lambda)$ [4]. A much more troublesome problem arises from the fact that luminous efficiency functions depend on the visual assessment.

Second, the additivity assumption may have failed. The integral in equation (1), the definition formula of the luminance, implies that the luminance of a stimulus composed by mixing stimuli of different wavelengths can be expressed by adding the luminance of the constituent monochromatic stimuli. In other words, it is assumed that the additivity law of luminance must be maintained. The additivity law is also referred to as Abney's law.

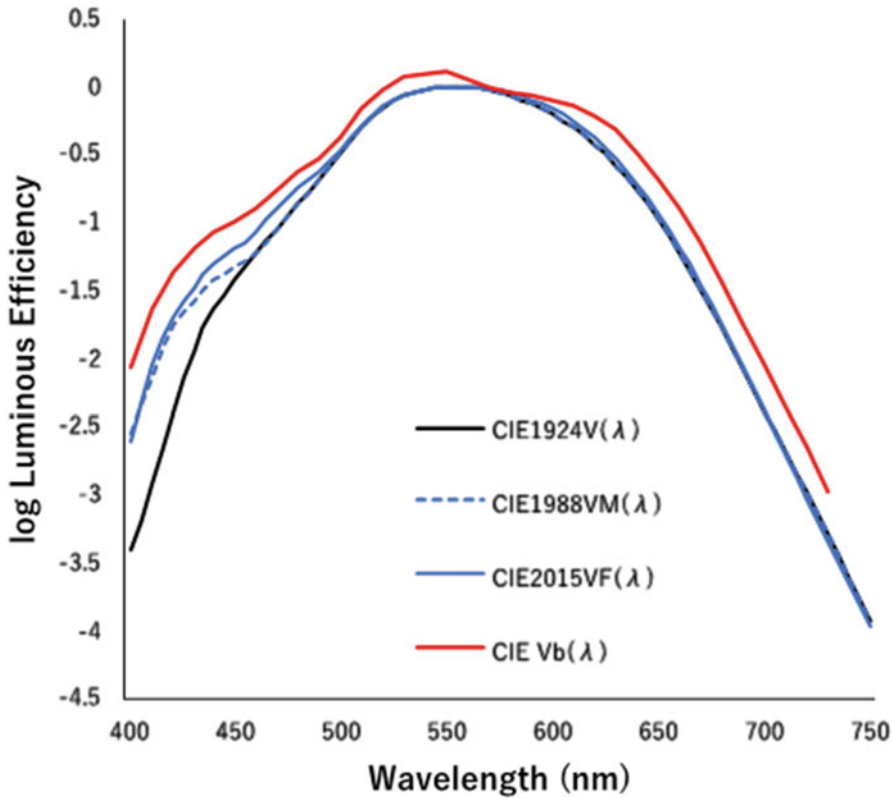
Luminous Efficiency Functions

A luminous efficiency function gives the ratio of the radiant power of a monochromatic stimulus of

wavelength λ_m to that of wavelength λ , when the two monochromatic stimuli produce luminous sensations that under given observing conditions and in accordance with some specified criterion are judged to be equivalent or the same, λ_m being chosen so that the maximum value of this ratio is unity.

The luminous efficiency function differs depending on the criteria embodied in the visual judgment. The major measurement procedures and criteria used in determining luminous efficiency functions include flicker photometry, direct heterochromatic brightness matching, step-by-step brightness matching, minimally distinct border, color matching, absolute threshold, increment threshold, visual acuity, and critical flicker frequency. The luminous efficiency functions by these methods are classified into two types. First, the flicker photometry is a typical example, and color does not contribute to the visual criteria. The 1924 CIE $V(\lambda)$ used in the definition of luminance was derived primarily from flicker photometry [5]. In the flicker photometry, the observer is asked to adjust the quantity of a chromatic light which temporally alternates with a reference light, until minimum flicker is obtained. The same type of flicker photometry was also obtained by the minimally distinct border, color matching, and critical flicker frequency methods. Even in heterochromatic brightness matching, when the visual field size is very small such as a point source is used, the luminous efficiency functions are close to the flicker photometry type.

On the other hand, the second type, which is typically the direct heterochromatic brightness matching, is affected by the difference in color on the judgment criteria. Since brightness matching is static, the criterion of brightness seems to be evaluated by not only the luminance component but also the chromatic component. There are many studies showing that the luminous efficiency function obtained by heterochromatic brightness matching has a broader curve than the CIE $V(\lambda)$ [6–8]. Based on early experimental data from these studies, the CIE provides a mean luminous efficiency function for heterochromatic brightness matching for 2° and 10° visual field [9]. Figure 1 shows the luminous efficiency



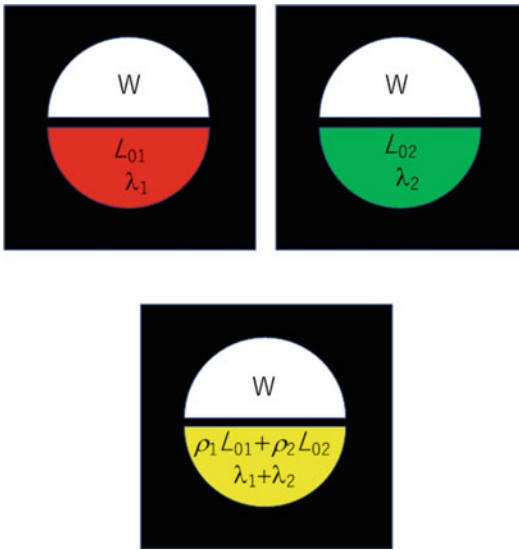
Helmholtz–Kohlrausch Effect, Fig. 1 Comparison of various luminous efficiency functions

function for heterochromatic brightness matching for 2° field, $V_{b,2}(\lambda)$ by comparing CIE1924 $V(\lambda)$, Judd's modification, CIE1988 $V_M(\lambda)$, and CIE2015 $V_F(\lambda)$ derived from the color matching functions [10]. $V_{b,2}(\lambda)$ shows higher efficiency than the other three luminous efficiency functions in the more saturated portion of the spectrum. It means that even if the luminance of each monochromatic light is set to be equal, the higher the saturation, the monochromatic light looks the brighter. The contribution of chromatic component to brightness is also based on another fact that the difference between the luminous efficiency function obtained by heterochromatic brightness matching and that obtained by flicker photometry is quite similar to the saturation-discrimination function obtained by measurements of just-noticeable colorimetric purity [8]. Since the saturation discrimination function is assumed to reflect the ratio of the quantity of activities of the chromatic channels to that of the

luminance channel, the difference between the two curves is considered to be related to the contribution of the chromatic components to brightness. The second type of luminous efficiency functions are also observed in the methods of the step-by-step brightness-matching, absolute threshold, and increment threshold methods.

Additivity Failures for Brightness

Luminance is defined on the assumption linear additivity of the contributions of different wavelengths as expressed by equation (1). Therefore, the visual criteria used for $V(\lambda)$ must obey the additivity law. The additivity law holds for minimum flicker which is the criterion in the flicker photometry used the basis for formulating the CIE 1924 $V(\lambda)$. The additivity law also holds for the criterion of minimally distinct border and for the criterion of visual acuity. Even in heterochromatic



Helmholtz–Kohlrausch Effect, Fig. 2 A schematic view of the additivity test for heterochromatic brightness matching

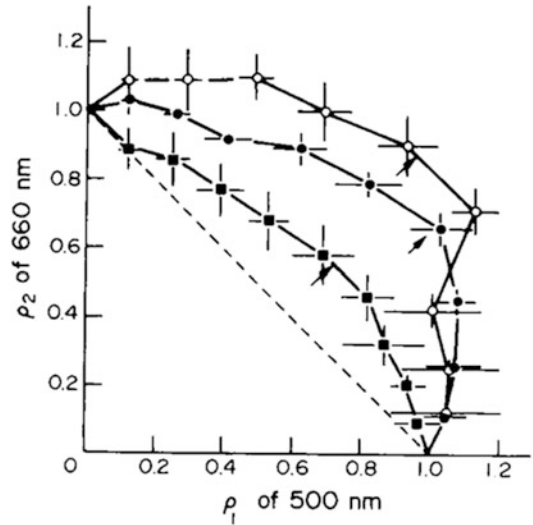
brightness matching, when a grating target or a fine checkerboard pattern is used, the additivity holds.

On the other hand, the additivity law is not obeyed with heterochromatic matching [7, 11]. Figure 2 shows a schematic view of a conventional additivity test of heterochromatic brightness matching. The radiances of monochromatic lights λ_1 and λ_2 be $L_{e,01}$ and $L_{e,02}$ that are matched to a white reference light in brightness, respectively, when they are presented separately. Next, monochromatic lights of these two wavelengths are mixed to match the brightness of the reference white light. The radiances of λ_1 and λ_2 be $L_{e,m1}$ and $L_{e,m2}$ are matched to the same white reference. The radiance ratios ρ_1 and ρ_2 are defined as follows:

$$\rho_1 = L_{e,m1}/L_{e,01} \quad (2)$$

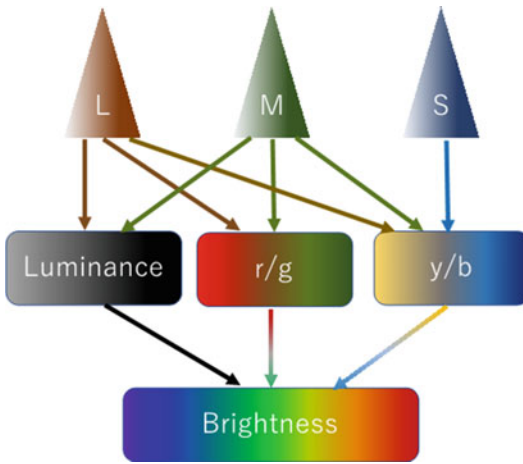
$$\rho_2 = L_{e,m2}/L_{e,02} \quad (3)$$

The additivity can be evaluated by the value, $\rho_1 + \rho_2$, such that if $\rho_1 + \rho_2 = 1$ the additivity holds, if $\rho_1 + \rho_2 > 1$ it fails as a reduction type or subadditivity, and if $\rho_1 + \rho_2 < 1$ it fails also but as an enhancement type or superadditivity.



Helmholtz–Kohlrausch Effect, Fig. 3 $\rho_1 - \rho_2$ plots for 500–660 nm pair from three subjects. Arrows indicate ratios of ρ_1 to ρ_2 of perceived red-green equilibrium for each subject

A typical plot of ρ_1 vs ρ_1 is shown in Fig. 3 where $\lambda_1 = 500$ nm and $\lambda_2 = 660$ nm for three subjects [11]. All points are located in the upper right region of a dashed straight line of linear additivity to indicate a reduction type additivity failure between the red and green lights, although its amount varies among subjects. The asymmetrical nature of the additivity failure is clearly seen in all subjects in that the largest deviation from the dashed line occurs not at the ratio $\rho_1/\rho_2 = 1$ but at a ratio larger than unity. The two colors employed here, named red and green, are complementary with each other and there must exist a ratio of ρ_1 to ρ_2 to give a red-green equilibrium, that is, neither reddish nor greenish. In the experiment, the subjects were asked to indicate the perceived hue of the mixture at various ratios. The arrows in Fig. 3 show the ratio at equilibrium. It is noticed in Fig. 3 that the maximum brightness reduction point exhibited by the largest deviation from the dashed line coincides with the arrow position of each subject. That is, the brightness reduction becomes strongest at the red-green equilibrium. Similar results were obtained with a mixture of yellow and blue lights. The results show a strong evidence that the chromatic components contribute to the brightness.



Helmholtz–Kohlrausch Effect, Fig. 4 A physiological color vision model for luminance and brightness

Physiological Mechanism and Model for Brightness

Figure 4 shows a physiological model of luminance and brightness based on the most reliable color vision model. It has been recognized that luminous quantity as measured by flicker photometry relies only on the sum of the activity of the long-wave sensitive (L-) and middle-wave sensitive (M-) cones. The signal of the difference between the L- and M-cones, L-M goes to the red-green opponent color channel and the L cone. The signal of the difference between the sum of L- and M-cones (L + M) and S-cones is transmitted to the yellow-blue opponent color channel. Luminous sensation in the flicker photometry that determines luminance is performed in the luminance channel without the contribution of the opponent color channel, and the brightness assessment in the heterochromatic brightness matching is made of two opponent color channels in addition to the luminance channel. As shown in Fig. 5, when a red light and a green light are mixed, the color components are canceled out in the red-green opponent color channel, causing the additivity failures of brightness. In this case, a unit brightness of red plus a unit brightness of green makes a unit brightness of yellow, that is, $1 + 1 = 1$. As a result, even with chromatic stimulus of the same luminance, more

colorful stimulus looks brighter than desaturated stimulus, that is, the Helmholtz-Kohlrausch effect.

Several mathematical models have been proposed that explain the mechanism of brightness. Among them, Guth et al. [7] proposed the vector model shown in equation (4).

$$L^{**} = (A^2 + T^2 + D^2)^{1/2}, \quad (4)$$

where the brightness L^{**} called vector luminance is represented by the vector sum of a non-opponent mechanism A , a red vs. green system T , and a yellow vs. blue system D . Ingling et al. [12] proposed similarly orthogonal combination of the three visual channels show in equation (5)

$$S^2 = Q^2 \left[(r - g)^2 + (y - b)^2 + V_\lambda^2 \right], \quad (5)$$

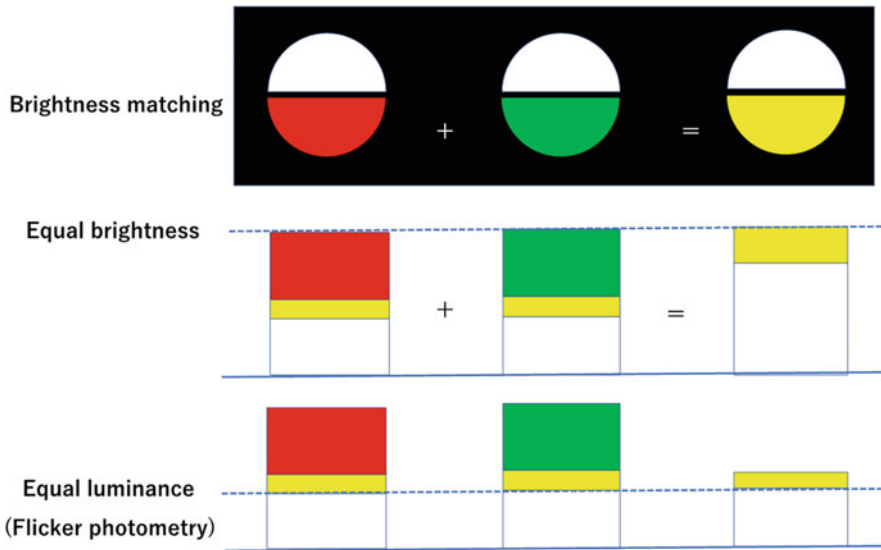
where S is the visual response such as brightness, Q is light intensity, $(r - g)$, $(y - b)$, and V_λ is the sensitivity of red-green opponent system, yellow-blue opponent system, and the luminance system, respectively. Yaguchi et al. [11] proposed a non-linear vector model by modifying the Guth's model to explain the asymmetrical additivity failure shown in Fig. 3. In their model, the luminous efficiency function for brightness $V_b(\lambda)$ is provided as equations (6):

$$\begin{aligned} (\bar{a}_\lambda L_{e,\lambda})^2 + \left(|\bar{c}_{1\lambda} L_{e,\lambda}|^{0.64} \right)^2 + \left(|\bar{c}_{2\lambda} L_{e,\lambda}|^{0.36} \right)^2 &= 1, \\ V_b(\lambda) &= 1/L_{e,\lambda}, \end{aligned} \quad (6)$$

where $L_{e,\lambda}$ is the radiance of a monochromatic light to provide a unit brightness, $\bar{a}_\lambda$, $\bar{c}_{1\lambda}$ and $\bar{c}_{2\lambda}$ are spectral sensitivity functions of the achromatic channel, the red-green opponent color channel, and the yellow-blue opponent color channel, respectively, defined for the equal energy spectrum.

Brightness-Luminance Ratio, B/L

The luminous efficiency functions deal only with monochromatic light. When a test stimuli of



Helmholtz–Kohlrausch Effect, Fig. 5 Relation between the Helmholtz-Kohlrausch effect and the additivity failures of brightness

complex spectral radiant power distribution has been matched in brightness against a white reference stimulus, it is generally found that the luminance of the chromatic test stimulus is lower than the luminance of the reference stimulus. The degree of the Helmholtz-Kohlrausch effect is usually evaluated by the ratio of the luminance of a reference stimulus, B to the luminance of the test stimulus, L , when they are matched in brightness, so called the B/L ratio. Many experimental data on B/L of the complex stimuli have been reported [13]. Ware and Cowan summarized the data sets from more than 50 studies including spectral and nonspectral lights, and provided an equation in which B/L is expressed as a fifth polynomial function of the CIE (x , y) chromaticity coordinates [14]. Several models have been proposed to predict the B/L ratio of complex stimuli [7, 11, 15].

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)

- [Helmholtz, Hermann Ludwig von](#)
- [Kohlrausch, Arnt](#)
- [Minimally Distinct Border](#)
- [Spectral Luminous Efficiency](#)

References

1. Wyszecki, G., Stiles, W.S.: Color Science. Wiley, New York (1982)
2. Judd, D.B.: A new look at the measurement of light and color. *Illum. Eng.* **53**, 61–71 (1958)
3. Judd, D.B.: Report of US. Secretariat Committee on Colorimetry and Artificial Daylight, CIE Proceeding, vol. 1, Part 7, p. 11, Stockholm (1951)
4. CIE 086-1990: CIE 1988 2° Spectral Luminous Efficiency Function for Photopic Vision. CIE Central Bureau, Vienna (1990)
5. Le Grand, Y.: Light, Colour and Vision. Chapman & Hall, London (1968)
6. Wagner, G., Boynton, R.M.: Comparison of four methods of heterochromatic photometry. *J. Opt. Soc. Am.* **62**, 1508–1515 (1972)
7. Guth, S.L., Lodge, H.R.: Heterochromatic additivity, foveal spectral sensitivity, and a new color model. *J. Opt. Soc. Am.* **63**, 450–462 (1973)
8. Yaguchi, H., Kawada, A., Shioiri, A., Miyake, Y.: Individual differences of the contribution of chromatic channels to brightness. *J. Opt. Soc. Am. A*, **10**, 1373–1379 (1993)

9. CIE 075-1988: Spectral Luminous Efficiency Functions Based Upon Brightness Matching for Monochromatic Point Sources 2° and 10° Fields. CIE Central Bureau, Vienna (1988)
10. CIE 170-2:2015: Fundamental Chromaticity Diagram with Physiological Axes – Part 2: Spectral Luminous Efficiency Functions and Chromaticity Diagrams. CIE Central Bureau, Vienna (2015)
11. Yaguchi, H., Ikeda, M.: Subadditivity and superadditivity in heterochromatic brightness matching. *Vision Res.* **23**, 1711–1718 (1983)
12. Ingling, C.R., Tsou, B.H.: Orthogonal combination of the three visual channels. *Vision Res.* **17**, 1075–1082 (1977)
13. Ayama, M., Ikeda, M.: Brightness-to-luminance ratio of colored light in the entire chromaticity diagram. *Color Res. Appl.* **23**, 274–287 (1998)
14. Ware, C., Cowan, W.B.: Specification of Heterochromatic Brightness Matches: A Conversion Factor for Calculating Luminance to Stimulus that Are Equal in Brightness. National Research Council of Canada Technical Report No. 26055. (1983)
15. Nayatani, N., Umemura, Y., Sobagaki, H., Takahama, K., Hashimoto, K.: Lightness perception of chromatic object colors. *Color Res. Appl.* **16**, 16–25 (1991)

Hering, Karl Ewald Konstantin

Rolf G. Kuehni
Charlotte, NC, USA

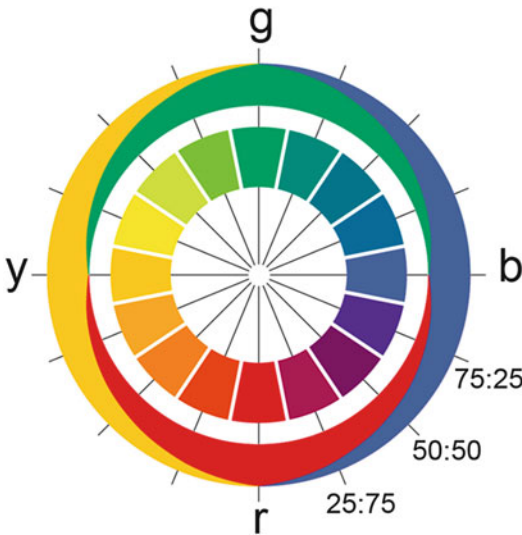


Biography

Karl Ewald Konstantin Hering was born on August 5, 1834, in Alt-Gersdorf, Saxony, son of a Lutheran pastor and his wife. He studied medicine at the University of Leipzig, obtaining an MD in 1860. For the next five years, he practiced medicine in Leipzig and pursued personal interests in vision on the side, publishing five *Beiträge zur Physiologie (Contributions to Physiology)* between 1861 and 1864 [1]. He was married in 1863, and he and his wife had a son, Heinrich Ewald, born in 1863. In 1865 he was appointed professor of physiology at the Josephinum Academy in Vienna. In 1870 he became the successor of J. E. Purkinje as professor of physiology at the University of Prague where he remained for 25 years, studying among other things the electrical actions of nerves and muscles, as well as the perception of light and color vision. In 1895 he was invited to join the University of Leipzig where he remained until his retirement in 1915. During his lifetime Hering assumed the role of anti-Helmholtz, scientifically and philosophically battling with him in regard to several subject matters [see, e.g., 2]. Hering died on January 26, 1918, in Leipzig.

Major Accomplishments/Contributions

Hering developed a theory, alternative in detail to that of Helmholtz, in regard to spatial perception based on images in two eyes. Publication of and the response to Helmholtz's first edition of the *Handbuch der Physiologischen Optik* in 1867 [3] provided a basis for Hering to consider in detail the issues of color perception and develop his own different theory. The result was the presentation in 1874 of six extended contributions to the Imperial Academy in Vienna, published in book format in 1878 as *Zur Lehre vom Lichtsinne (On the Theory of the Light Sense)* [4]. Hering considered the Young-Helmholtz theory to give too much weight to physics and not enough to the perceptual aspects of color. In response he developed an opponent-color theory postulating three



Hering, Karl Ewald Konstantin, Fig. 1 Superimposed images of the conceptual mixture of the four hue fundamentals in different ratios and the resulting hue circle

opponent pairs of fundamental colors (*Urfarben*), yellow-blue, red-green, and white-black, building on earlier ideas by H. Aubert and E. Mach. The hues are arranged according to simple principles in a hue circle (Fig. 1) (derived from two images in Ref. [4]. Image Tilo Hauke), the complete arrangement of all colors of a given hue being in a triangle, with white, black, and full color at the corners and “veiled” colors filling the interior. Hering called the result, in form of a double cone, the “natural color system.” He also proposed dissimilation/assimilation processes in the eye/brain as physiological basis of the opponent system, in opposition to the Young-Helmholtz theory involving three fundamentals. A commercial version of the Natural Color System atlas was introduced in 1979 by the Scandinavian Colour Institute [5].

Today, the perceptual aspects of Hering’s system continue to be considered essentially valid, with its physiological part having fallen short of reality. Despite many attempts, a psychophysical model of the four hues presumed fundamental and their mixtures that also meet other components of the colorimetric system is still lacking at this time, a major issue being the fact that while mean unique yellow and blue stimuli are essentially

complementary, unique green and unique red stimuli are far from it, complicating any model with the purpose of representing a perceptually meaningful and at the same time colorimetrically valid model.

References

1. Hering, E.: Beiträge zur Physiologie. Engelmann, Leipzig (1861)
2. Turner, R.S.: In the Eye’s Mind, Vision and the Helmholtz-Hering Controversy. Princeton University Press, Princeton (1994)
3. von Helmholtz, H.: Handbuch der physiologischen Optik. Voss, Leipzig (1867)
4. Hering, E.: Zur Lehre vom Lichtsinne. Gerold, Wien (1878). Translation: Outlines of a theory of the light sense. Hurvich, L. M., Jameson, D., trans. Harvard University Press, Cambridge, MA (1964)
5. Scandinavian Colour Institute: Swedish Natural Colour System (NCS)

Hertzsprung-Russell Diagram

Michael H. Brill

Datacolor, Lawrenceville, NJ, USA

Definition

A scatter graph of stars showing the relationship between each star’s absolute magnitude and its estimated surface temperature, or between optical and perceptual proxies for these quantities.

Absolute Magnitude and Temperature Scales

The absolute magnitude and temperature can be described as follows:

A star’s absolute magnitude is the attenuation (in factors $10^{-0.4}$) of the star Vega’s power (as received at 10 pc viewing distance [32.6 light-years]) to equal that of the star (also corrected to 10 pc). The convention that dimmer stars have higher magnitude is a historical

precedent that dates from Hipparchus (c. 190 BCE–c. 120 BCE), whose system of stellar magnitudes was based on visual assessment.

A star's surface temperature is estimated in one of three ways: by the observed color (an old way), by comparison of two sensor outputs such as blue and violet (a newer way), or by model prediction of the temperature of a black-body radiator with the same radiation power per unit star-surface area (the most modern way). The third way requires independent inference of the star's radius but assumes the star has an emissivity of 1. To acknowledge the lack of compensation for the true emissivity, the temperature on an H-R diagram is called "effective temperature."

Coordinate Conventions

Here are the coordinate conventions of the H-R diagram: Temperature increases from right to left, and magnitude (i.e., dimness) increases from bottom to top. Hence, dim red stars (red giants) appear near the upper right of the diagram, and bluish bright stars (such as white dwarfs) appear in the lower left. A long cluster of stars called the main sequence extends from the upper left to the lower right. Higher mass stars occur at the upper left of this sequence, and the Sun appears approximately in the middle. The main sequence is composed of stars that are currently dominated by hydrogen that is fusing into helium. According to currently accepted theory of stellar evolution, such stars will eventually migrate either to the red giants or white dwarf domain of the H-R diagram.

H-R diagrams are often depicted in color, either pseudocolor with a thermal code to show the temperature or coded according to star categories such as cluster membership.

History

The H-R diagram was originated by Danish astronomer Ejnar Hertzsprung and American astronomer Henry Norris Russell (https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram). Originally the diagram was based on visual

estimation of magnitude and color, and it was a research tool to help characterize stellar evolution before the mechanism of nuclear fusion was understood. After about the 1930s, the H-R diagram became based on objective measurements but was used less as a research tool and more as a way to illustrate the theoretically predicted evolution of stars through trajectories in the diagram.

Cross-References

- [Color Magnitude Diagrams](#)
- [Galaxy Color Magnitude Diagram](#)

High- and Low-Pressure Sodium Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Sodium gas discharge lamps](#)

Definition

A lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a sodium vapor that is contained in a transparent bulb.

Sodium Gas Discharge Lamps

Sodium gas discharge lamps can be distinguished in two main groups: low-pressure sodium lamps

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, the Netherlands. © 2012 Koninklijke Philips Electronics N.V.

in which the gas pressure is very low (0.7 Pa or 7.10^{-6} atm) and high-pressure sodium lamps in which the gas pressure is a factor 10,000–100,000 times higher (10–100 kPa) [2–4]. Because their construction and performance are in many ways different, they are dealt with in two different sections of this entry.

Low-Pressure Sodium Lamps

Low-pressure sodium lamps belong to the group of high-intensity discharge (HID) lamps because they are available in high-light output (and thus high-luminous-intensity) versions. All low-pressure gas discharge lamps have in common the fact that they are long. Low-pressure sodium lamps are highly efficient lamps with a good lifetime but no color rendition at all. Their application is therefore restricted to those situations where color rendering is of no importance, as, for example, on motorways, on railway-marshaling yards, and in some security-lighting situations. Low-pressure sodium lamps are sometimes also referred to as LPS lamps.

Working Principle

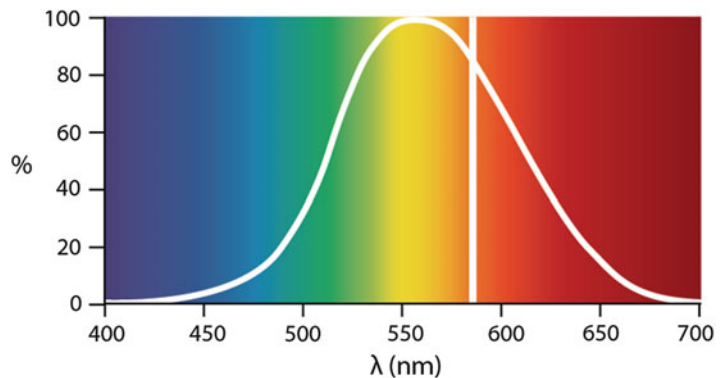
The gas discharge principle of low-pressure sodium lamps is similar to that of low-pressure mercury lamps (see that chapter for general details). In low-pressure sodium lamps, the discharge takes place in vaporized sodium. The low-pressure sodium discharge emits monochromatic radiation in the visible range. Therefore, unlike low-pressure mercury lamps, they do not need fluorescent powders to convert the wavelength of the radiation. The monochromatic (single wavelength) radiation is the reason that color rendering is nonexistent. The wavelength of the monochromatic radiation is 589 nm (yellowish light), which is very close to the wavelength for which the eye has its maximum sensitivity (Fig. 1). It is mainly for this reason that the lamp has such an extremely high-luminous efficacy (up to 190 lm/W system efficacy). Like all gas discharge lamps (with very few exceptions), a low-pressure sodium lamp cannot be operated without a ballast to limit the current flowing through it.

Materials and Lamp Construction

The main parts of a low-pressure sodium lamp are (Fig. 2):

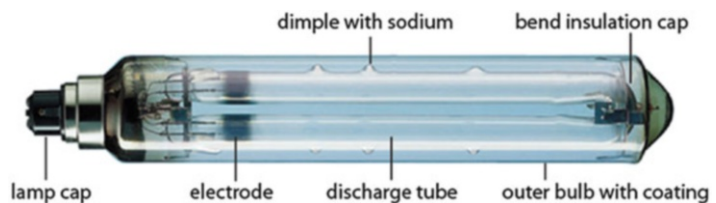
High- and Low-Pressure Sodium Lamp,

Fig. 1 Relative eye-sensitivity curve and the monochromatic line of 589 nm of the low-pressure sodium spectrum



High- and Low-Pressure Sodium Lamp,

Fig. 2 Principle parts of a low-pressure sodium-discharge lamp [1]



- Discharge tube
- Fill gas
- Electrodes
- Outer bulb with inner coating
- Lamp cap

Discharge Tube

Just as with the low-pressure mercury, fluorescent tube, and lamp, the power dissipated in the low-pressure sodium lamp largely determines the length of the discharge tube. Especially for the higher wattages (and thus lumen packages), the unfolded length has to be really long. To reduce the actual length of the lamp, the discharge tube of low-pressure sodium lamps is therefore always U shaped. Nevertheless, the highest lumen packages still require a lamp length of about 1.2 m. The U-shaped discharge tube is made of sodium-resistant glass and contains a number of small dimples, or hollows, where the sodium is deposited as a liquid during manufacture. After ignition, the discharge first takes place through the inert gas mixture. As the temperature in the tube gradually increases, some of the sodium in the dimples vaporizes and takes over the discharge, which then emits the monochromatic radiation. At switch-off, the sodium condenses and again collects at the dimples, these being the coldest spots in the tube. Without the dimples, the sodium would, after some switch-on switch-off cycles, gradually condense along the whole inner tube wall, decreasing light transmission considerably.

Fill Gas

The inert gas mixture of neon and argon, called the “Penning mixture,” acts as a starting gas and buffer gas (to protect the electrodes). During start-up, the discharge only takes place in this gas, which is why a low-pressure sodium lamp radiates deep-red light for some 10 min during start-up (Fig. 3).

Electrodes

Most modern low-pressure sodium lamps have cold-start electrodes. These consist of a triple-coiled tungsten wire so that they can hold a large quantity of emitter material.

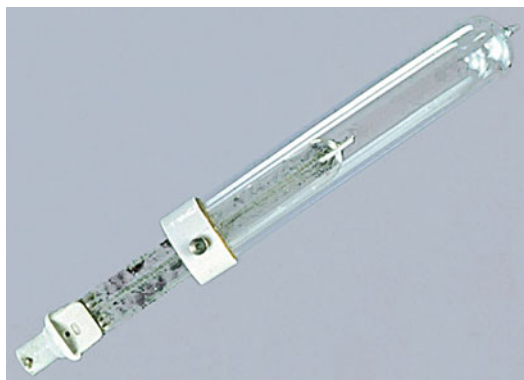


High- and Low-Pressure Sodium Lamp, Fig. 3 Immediately after switch-on, a low-pressure sodium lamp emits just a little reddish light, which gradually changes to the familiar yellow sodium light when the lamp has fully warmed up [1]

Outer Bulb

The optimum sodium-vapor pressure is reached when the temperature of the wall of the discharge tube is maintained at 260 °C. To reach this temperature efficiently, the U-shaped discharge tube is contained in an evacuated outer-glass tube. To further increase the thermal insulation, this outer-glass tube is coated on its inner surface with an interference layer that reflects infrared radiation but transmits visible radiation. In this way, most of the heat radiation is reflected back into the discharge tube, so maintaining the tube at the desired temperature, while visible radiation is transmitted through the layer. Early low-pressure sodium lamps did not have an integrated outer bulb with infrared reflecting coating. They used a separate double-wall evacuated tube (“thermos bottle”) which was reused after the failure of the gas discharge tube at the end of its life (Fig. 4).

Apart from the dimples, the bend of the U-shaped discharge tube also forms a cold spot



High- and Low-Pressure Sodium Lamp, Fig. 4 Immediately after switch-on, a low-pressure sodium lamp emits just a little reddish light, which gradually changes to the familiar yellow sodium light when the lamp has fully warmed up

where sodium can condense and accumulate. The bend is therefore insulated by a heat-reflecting metal cap.

Lamp Cap

All low-pressure sodium lamps are provided with a bayonet-type lamp cap. This allows the discharge tube to be accurately positioned. This is critical because the light distribution of a low-pressure sodium luminaire is dependent on the position of the U-shaped discharge tube.

Properties

Energy Balance

A low-pressure sodium lamp emits approximately 40% of the input power in the form of visible radiation. This is the highest percentage of all gas discharge lamps. The remaining part of the input power is lost in the form of heat.

System Luminous Efficacy

The luminous efficacy of the system is strongly dependent on the wattage of the lamp and ranges from 70 to 190 lm/W, for low and high wattages, respectively.

Lumen-Package Range

Low-pressure sodium lamps are available in the range from approximately 2,000 to 30,000 lm (corresponding wattage range, 18–180 W).

Color Characteristics

As mentioned before, low-pressure sodium lamps emit monochromatic light in the yellowish part of the spectrum (Fig. 5). Color rendering is therefore nonexistent ($R_a = 0$). The correlated color temperature is around 1,700 K.

Lamp Life

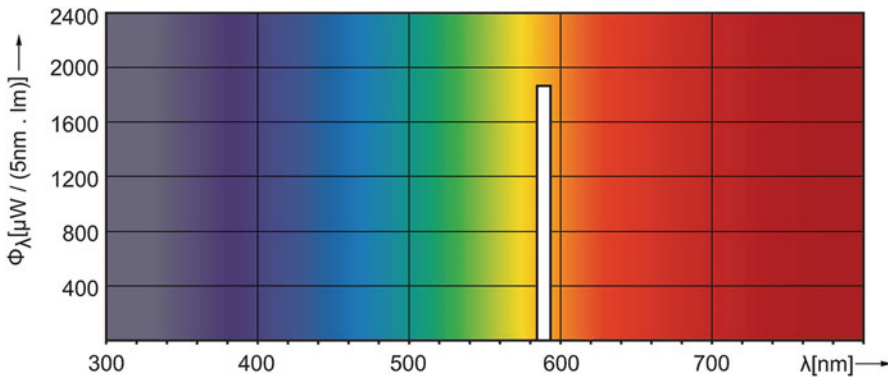
Apart from the usual cause of failure in gas discharge lamps (viz., electrode emitter exhaustion), low-pressure sodium lamps may also fail because of cracks or leaks in the long discharge tube or outer bulb. This may especially be the case in environments where there are strong vibrations as, for example, may occur in poorly designed road-lighting luminaires during strong winds. Leakage in the outer bulb disrupts the thermal isolation, which in turn means that not enough sodium will vaporize. As a consequence, the discharge will take place in the starting-gas mixture, so emitting only the corresponding deep-red light. Economic lamp life is around 12,000 h (based on a 20% mortality rate). Some versions make use of special getter material to maintain a high vacuum, resulting in fewer failures during the economic life of the lamp: the economic lamp lifetime is about 15,000 h (again, mortality rate of 20%).

Lamp-Lumen Depreciation

Lumen depreciation occurs through blackening of the discharge tube by scattering of the emitter material of the electrodes and by discoloration of the glass caused by the sodium. Depending on the type of control gear used, these effects are partially counteracted by a slow and gradual increase in the power dissipated in the lamp.

Burning Position

Electrodes and lead-in wires coming into contact with condensed sodium can eventually suffer damage. To prevent this from happening, low-pressure sodium lamps have restrictions as to their burning position. Base-down burning positions, in particular, have to be avoided. The restrictions are less critical for lower-wattage (viz., smaller) lamps, simply because they contain less sodium.



High- and Low-Pressure Sodium Lamp, Fig. 5 Spectral energy distribution of low-pressure sodium lamps [1]

Run-Up and Reignition

As has already been explained, the sodium needs time to vaporize while the discharge takes place in the starting-gas mixture. The warming-up process takes about 10 min. Nearly all low-pressure sodium lamps reignite immediately. The exceptions are the highest wattage lamps (131 and 180 W), which restrike after 10 min.

Dimming

Low-pressure sodium lamps cannot be dimmed. Dimming would decrease the lamp temperature so that not enough sodium would remain in the vapor state to maintain the sodium discharge.

Ambient-Temperature Sensitivity

The excellent thermal insulation afforded by the outer bulb ensures that lamp performance is almost independent of ambient temperature. Also, thanks to the starting-gas mixture, starting too is almost independent of ambient temperature.

Mains-Voltage Variations

The variations in lamp current and lamp voltage as a consequence of a change in mains supply voltage tend to cancel each other out, the net result being that the lamp wattage and, to a certain extent, the luminous flux remain practically constant over a wide range.

Product Range

Low-pressure sodium lamps are available in two versions, each with a different balance between

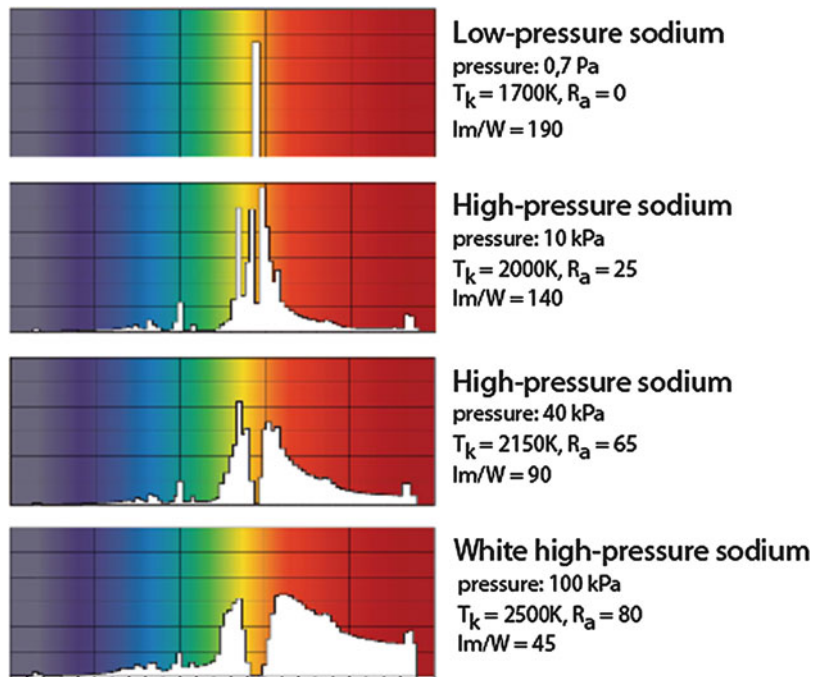
efficacy and lumen output. There is a high-lumen-output version with a 10–15% lower efficacy and a high-efficacy version with a 20% lower lumen output, respectively. The high-efficacy lamps can be operated on HF electronic control gear, which further increases their efficacy by between 15% and 35%. The length of low-pressure sodium lamp increases considerably with wattage: The 18 W version (the lowest wattage available) has a length of 22 cm, while the longest lamps (131 and 180 W versions) have a length of 112 cm.

High-Pressure Sodium Lamps

High-pressure sodium gas discharge lamps as well as low-pressure sodium lamps belong to the group of HID lamps because they are available in high-light output (and thus high-luminous-intensity) versions. High-pressure sodium lamps are also referred to as HPS lamps. High-pressure sodium lamps, in common with all high-pressure discharge lamps, are relatively compact. By increasing the vapor pressure in a sodium lamp, the color rendering improves and the color appearance changes from yellow to yellow-white, albeit at the cost of a decrease in efficacy. However, the resulting efficacy is more than double that of a high-pressure mercury lamp. At its introduction in the late 1960s, a very efficient alternative was thus obtained for the many high-pressure mercury lamps employed at that time in

High- and Low-Pressure Sodium Lamp,

Fig. 6 Effect of sodium-vapor pressure on the spectral power distribution of different sodium gas discharge lamps [1]



H

road lighting. Today, road-lighting installations all over the world very often use high-pressure sodium lamps, although for some installations LED solutions have become an alternative.

By further increasing the sodium pressure, the color quality of the light improves to such an extent that the light becomes real white. These so-called white high-pressure sodium lamps have a lower efficacy but sometimes offer an acceptable alternative to halogen and compact metal halide lamps for accent lighting.

Working Principle

It has been shown that with low-pressure sodium lamps at the low working pressure of that discharge, a single, monochromatic line of light at a wavelength of 589 nm is emitted. With increasing pressure, the radiation in the core of the discharge is absorbed by the cooler surrounding gas and reemitted in the form of radiation not of the 589 nm line but with wavelengths slightly smaller and slightly larger than 589 nm. So, the 589 nm line gradually disappears (called self-absorption), while in the wavelength area to the left and right of that value, more and more light is emitted (broadening of the spectrum). The

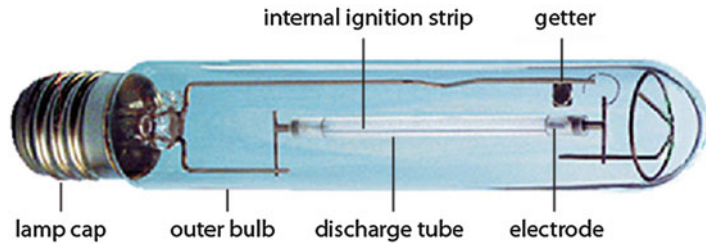
phenomenon of self-absorption and spectrum broadening is illustrated in Fig. 6, which shows examples of sodium lamps with different operating vapor pressures. The phenomenon is accompanied by a loss of efficacy. At the operating pressure of a standard high-pressure sodium lamp, the lamp has a yellow-white color appearance (2,000 K) and a moderate color-rendering index R_a of approximately 25 at an efficacy, for the higher wattages, of some 140 lm/W. With further increase in operating pressure, the same process continues, viz., widening of the spectrum at the cost of efficacy. With lamps that operate at a four-times-higher pressure, the color rendering improves to “fairly good” ($R_a = 65$) at an efficacy of around 90 lm/W. A version with an operating pressure ten times higher than that of the standard high-pressure sodium lamp is also produced: white HPS. This version has an R_a of 80 and radiates white light with a color temperature of 2,500 K at an efficacy of around 45 lm/W.

Materials and Construction

The main parts of a high-pressure sodium lamp are (Fig. 7):

High- and Low-Pressure Sodium Lamp,

Fig. 7 Main parts of a high-pressure sodium lamp (e.g., tubular version) [1]



- Discharge tube
- Fill gas
- Outer bulb
- Electrodes
- Lamp caps
- Getter

Discharge Tube

Sodium at high temperatures has the tendency to migrate slowly through quartz. Quartz can therefore not be used as material for discharge tubes of high-pressure sodium lamps. Discharge tubes made of ceramic material withstand sodium at very high operating temperatures. This is why, right from their introduction in 1966, high-pressure sodium lamps have translucent, tubular-shaped, ceramic discharge tubes. Since the ceramic tube is produced by sintering minuscule aluminum-oxide particles, one cannot look straight through the wall, although the light transmittance is more than 90% (translucent instead of transparent material).

Fill Gas

Sodium is introduced into the gas discharge tube as a sodium-mercury amalgam composition, which partially vaporizes when the lamp reaches its operating temperature. The sodium vapor is responsible for excitation and subsequent light radiation, while the mercury gas acts to regulate the voltage of the lamp and reduces thermal losses (buffer gas). The starter gas commonly added is xenon. An exception to this is found in the case of low-wattage lamps, which have a built-in igniter in the form of a bimetal switch and a neon-argon mixture – the Penning mixture also used in fluorescent lamps – as starting gas. By increasing the pressure of the starting-gas xenon, the luminous efficacy of the lamp increases by about 20%.

However, to ensure proper ignition at this higher starting-gas pressure, an auxiliary ignition wire has to be added very close to the discharge tube. Some versions of high-pressure sodium lamps, therefore, have an internal ignition strip (see Fig. 7). From an environmental point of view, it can be desirable to have high-pressure sodium lamps that do not contain mercury. Such lamps are, in fact, available. In these mercury-free high-pressure sodium lamps, xenon is used not only as a starting gas but also as the buffer gas that regulates the voltage and reduces thermal losses.

Electrodes

The electrodes employed in high-pressure sodium lamps are basically the same as those found in high-pressure mercury lamps. They consist of a core of tungsten rod with a tungsten coil (impregnated with emissive material) wound around it.

Outer Bulb

To thermally insulate the gas discharge tube and to protect its components from oxidation, an outer bulb is employed. The outer bulb of standard high-pressure sodium lamps is either ovoid or tubular (T version) in shape (Fig. 8). The internal wall of the ovoid bulb is usually coated with a diffusing powder. The coated versions were introduced so as to obtain the same light-emitting area as in regular ovoid, fluorescent-powder-coated, high-pressure mercury lamps. In this way, the coated ovoid high-pressure sodium lamps can be used with the same luminaire optics as those developed for high-pressure mercury lamps. This was especially important at the original introduction of high-pressure sodium lamps when they were replacing many existing high-pressure mercury lamps that were then a standard in many

High- and Low-Pressure Sodium Lamp,

Fig. 8 Different outer bulbs and lamp caps of high-pressure sodium lamps



road-lighting installations. Note that the coating in high-pressure sodium lamps is of the diffusing, nonfluorescent type. Since high-pressure sodium lamps produce practically no UV radiation, there is no point in using fluorescent powder. White high-pressure sodium lamps are usually only available in the tubular form. The glass used for the outer bulb of high-pressure sodium lamps for wattages of more than 100 W is hard glass.

Lamp Caps

The lamp caps employed for standard high-pressure sodium lamps are of the Edison screw type. The white high-pressure sodium lamps have a special bi-pin cap to ensure exact positioning in a luminaire.

Getter

During the process in which the lamp is evacuated, it is impossible to remove all traces of air and water vapor. During the operation of the lamp, minuscule particles evaporate from the glass and metals in the tube. All these traces would lead to an unacceptably short life. To remove them a getter is added that absorbs these traces. The getter is usually in the form of a small piece of solid material (see Fig. 7).

Properties

Energy Balance

Some 30% of the input power of a middle-range type of high-pressure sodium lamp is emitted in the form of visible radiation. Compare this with

the 40% of low-pressure sodium lamps and 17% of high-pressure mercury lamps.

System Luminous Efficacy

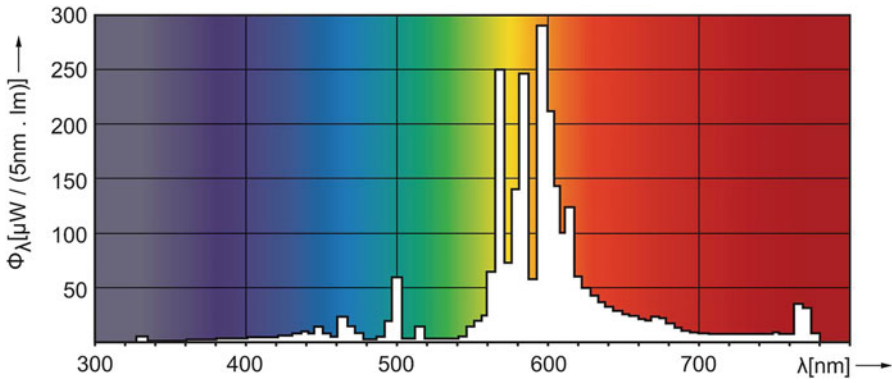
The efficacy of the compact white high-pressure sodium lamp varies between approximately 30 and 45 lm/W. The high-pressure sodium lamp with a color-rendering index of 60 has an efficacy between 75 and 90 lm/W, and the standard high-pressure sodium lamps (with R_a around 25) has an efficacy between some 80 and 140 lm/W. The higher the wattage, the higher the efficacy.

Lumen-Package Range

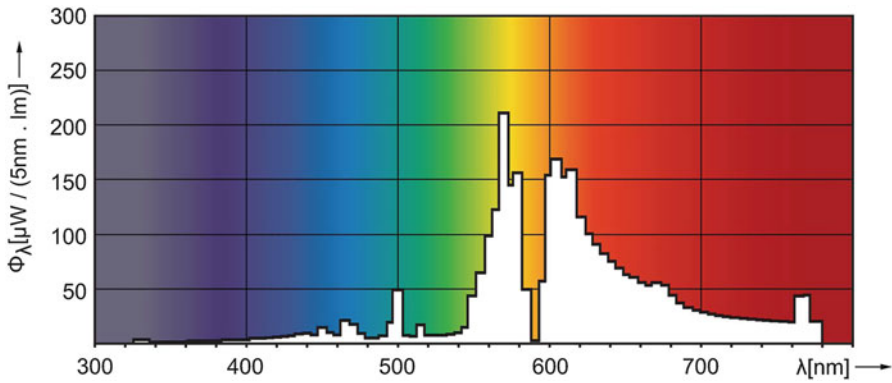
Compact white high-pressure sodium lamps are available in lumen packages from some 1,500 to 5,000 lm (corresponding wattage range of 35–100 W). Standard high-pressure sodium lamps are being produced in the approximate range of 4,000–150,000 lm (corresponding wattage range of 50–1,000 W).

Color Characteristics

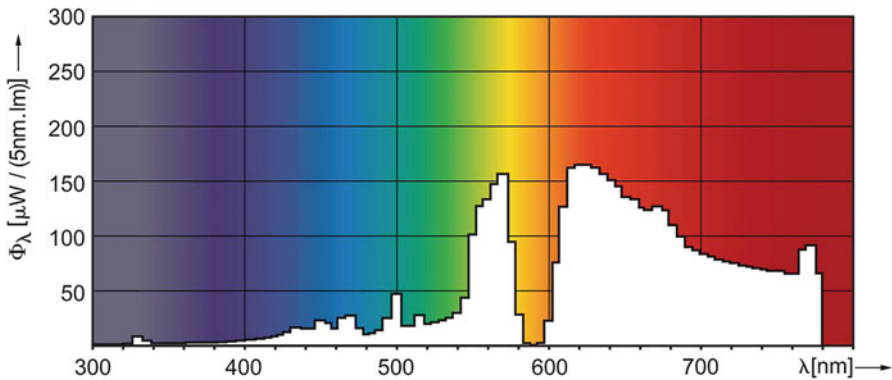
As with all gas discharge lamps, the high-pressure sodium lamp spectrum is discontinuous. Figures 9, 10, and 11 show the spectra of a standard high-pressure sodium lamp, a color improved lamp, and a white lamp. The color temperature of these versions ranges from 2,000 to 2,500 K and the color-rendering index from 25 to 80. Since the spectrum of all versions is relatively strong in the red wavelength area, the rendition of human faces is often experienced as being somewhat flattering. Of course, for indoor lighting, the



High- and Low-Pressure Sodium Lamp, Fig. 9 Spectral energy distribution of a standard high-pressure sodium lamp: T_k 2,000 K and R_a 25 [1]



High- and Low-Pressure Sodium Lamp, Fig. 10 Spectral energy distribution of a color improved high-pressure sodium lamp: T_k 2,150 K and R_a 65 [1]



High- and Low-Pressure Sodium Lamp, Fig. 11 Spectral energy distribution of a white high-pressure sodium lamp: T_k 2,500 K and R_a 80 [1]

color rendering of the standard high-pressure sodium lamp is far from adequate. For road lighting, it is experienced as being acceptable.

Lamp Life

The lamp voltage of a high-pressure sodium lamp increases gradually with life. The chief cause of lamp failure is that the lamp voltage rises higher than the voltage output of the ballast, causing the lamp to extinguish. When this happens, the lamp cools down, and the pressure in the lamp decreases so that the igniter can ignite the lamp again. After some minutes, the lamp voltage again increases too much, and the lamp extinguishes again. So, the usual end of life of a high-pressure sodium lamp is accompanied by this so-called cycling effect. For those situations where this on-off cycling might be disturbing or would damage the ballast, special “self-stopping” gear is available that stops igniting the lamp once the high voltage at the end of life is reached.

Standard high-pressure sodium lamps have an economic life of up to some 20,000 h (20% mortality). The lifetime of white high-pressure sodium lamps is not determined by the moment of actual failure of the lamps but by the onset of too large a color shift of the light, which is caused by the gradual increase of lamp voltage. To substantially increase their lifetime, white high-pressure sodium lamps, therefore, employ an electronic voltage stabilizer integrated into their control gear. The economic lifetime of compact white high-pressure sodium lamps, depending on type, lies between some 8,000 and 12,000 h (based on a 20% too-large color shift).

Lamp-Lumen Depreciation

For the compact white lamps, lumen-depreciation values vary between some 20% and 25% (after 10,000 h). The standard high-pressure sodium lamps have a much smaller lumen depreciation of between approximately 5% and 10% (after 20,000 h).

Run-Up and Reignition

The high-pressure sodium lamp must be ignited by a high-voltage pulse, typically 1.8–5 kV. After

ignition, the color of the light is initially white (discharge in the starting gas), changing to yellowish after some 20 s as the sodium amalgam gradually vaporizes and the vapor pressure rises; until after some 3–5 min, the nominal pressure and full light output are reached. The reignition of the hot lamp requires the lamp to cool down for about 1 min to allow the pressure to decrease to a point where the ignition pulse can again ionize the sodium atoms.

Dimming

All high-pressure sodium lamps can be dimmed to a certain extent, depending on the type of dimming equipment used. Lower wattages (100–150 W) can be dimmed with special electronic gear, which allows for dimming to 20%. Higher lamp wattages can be dimmed by including an extra inductive coil (ballast) in the ballast circuit. Lamp color remains virtually constant and lifetime is not affected.

Ambient-Temperature Sensitivity

The behavior of high-pressure sodium lamps during temperature variations differs from that of other discharge lamps because of the excess of amalgam used in the lamp. Although the outer bulb offers some degree of thermal isolation, the lamp manufacturer's specifications should be followed in the design of luminaires as far as its effect on lamp temperature is concerned.

Mains-Voltage Variation

A 5% mains-voltage variation has a 15% effect on the light output of high-pressure sodium lamps. The same 5% mains-voltage variation has a 5% effect on lamp voltage.

Product Range

As shown, high-pressure sodium lamps are available in three basic types: standard high-pressure sodium, an improved color version of high-pressure sodium, and a compact, white-light, high-pressure sodium lamp. The standard high-pressure sodium lamp is available in two forms: the standard one and an extra-high-efficacy version (with internal ignition strip). Most high-

pressure sodium lamps contain a tiny quantity of mercury, but there is also a version that is entirely free of mercury.

For ease of replacement of high-pressure mercury lamps in existing installations with more efficient high-pressure sodium types, a special type of high-pressure sodium lamp has been developed. This type uses a neon-argon (Penning) mixture as starting gas and is fitted with an ignition coil surrounding the discharge tube. These features allow the lamp to be operated on a standard high-pressure mercury-lamp ballast.

Cross-References

- [High Pressure Mercury Lamp](#)
- [Metal Halide Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Coursebook. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: *Lamps and Lighting*, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *The Lighting Handbook Tenth Edition: Reference and Application*. Illuminating Engineering Society of North America (IESNA), New York (2011)
4. De Groot, J., Van Vliet, J.: *The High-Pressure Sodium Lamp*. Kluwer Technische Boeken, Deventer/Antwerpen (1986)

High Dynamic Range Imaging

Tania Pouli
Advanced Media Contents Lab, IRT b-com,
Cesson-Sévigné, France

Synonyms

[Extended dynamic range](#); [HDR](#); [HDRI](#); [Wide-dynamic-range imaging](#)

Definition

High dynamic range imaging (HDRI or HDR) is a collection of hardware and software technologies that allow the capture, processing, and display of image and video content containing a wide range of intensities between the darkest and brightest areas of an image. Lighting conditions in both natural and man-made environments can range from starlight to artificial illumination to bright sunlight. Traditional imaging techniques typically store information using 8 bits per pixel for each channel, allowing for 256 distinct steps per channel. Consequently, they can only represent a narrow range of this illumination, often resulting in over- or undersaturated regions in the image. In contrast, HDR technologies represent image content using a higher precision (10 or 12 bits, or floating-point numbers) and thus both allow for a much larger number of intensity steps to be encoded, reducing the distance between consecutive steps in image intensity.

Overview

Natural environments contain a vast range of illumination that extends from bright sunlight to starlight. In fact, lighting conditions in natural scenes can extend over more than 14 orders of magnitude. Human eyes are not able to perceive this enormous range simultaneously, but mechanisms in the visual system allow adaptation to the prevailing lighting conditions and surroundings. A recent study has evaluated the simultaneous range of the human visual system under extended luminance levels and has found that under specific circumstances, it can reach (and theoretically exceed) 3.7 orders of magnitude [1].

Traditionally, digital images are stored using 8 bits per pixel to represent the values for each of the three color channels, allowing for 256 distinct levels between the darkest and brightest value within each channel. Consequently, standard imaging techniques can capture and store roughly 2 orders of magnitude between black and white, separated by distinct steps. Any information exceeding the available range appears over- or

underexposed, while values between intensity levels are mapped to one of the 256 available levels, introducing steps into smooth gradients – a process known as quantization.

This is a familiar problem for photographers, as many cases exist where this range cannot adequately capture the full scene. For instance, when capturing the interior of a room, if there is a window with bright sunlight outdoors, the region covered by that window would be significantly brighter than the interior. The photographer would then have to choose whether to expose for the room interior or the window to be captured correctly, losing the remaining information of the scene. Such a solution could be sufficient in an artistic context; however, when images are intended as a realistic and complete representation of a scene, the photographer's choices and limitations of the equipment can affect the results. Useful information may not be captured or images may include bias, as the photographer needs to select an exposure level. Figure 1 shows two photographs of the same scene exemplifying this problem.

High dynamic range (HDR) imaging is a collection of tools and algorithms that allow the capture, storage, processing, and display of images with a wider dynamic range than that afforded by traditional imaging techniques. High dynamic range images and video content use a higher precision to encode luminance information

relative to the 8-bits per color channel traditionally used in previous formats, collectively referred to as standard dynamic range (SDR). This affords two key advantages: first, a much wider range of illumination information can be encoded. Second, the higher precision used (which can vary from 10- or 12-bits per color channel in video formats to floating point number in image formats) means that HDR images have significantly less quantization.

Imaging Pipeline

The inherent differences between HDR and traditional (SDR) imagery, in addition to allowing more information to be captured and opening up many possibilities in terms of applications, also mean that all stages of an imaging pipeline need to be modified. From the capture of HDR images and video, to encoding and high dynamic range storage solutions, to processing, and finally to display, many existing algorithms make assumptions that do not apply to images and video content with a wider dynamic range.

Capture

To capture the full dynamic range of a scene such as the one shown in Fig. 1, a camera would need to be able to capture both bright and darker areas simultaneously. To achieve this, several solutions



High Dynamic Range Imaging, Fig. 1 Both photographs capture the same scene but using different exposure levels. The foliage and colors of the tree are well exposed in the *left image* but the sky has been overexposed and has no details visible. The *right image* on the other hand shows

the details of the cloudy sky much better but at the expense of most information in the rest of the image. Problems such as these can easily be overcome with the use of high dynamic range technologies



High Dynamic Range Imaging, Fig. 2 An HDR image (here shown tonemapped) constructed by combining the seven exposures shown on the *left*. Each individual exposure cannot fully capture all the detail in the scene

have been proposed. A common approach suitable for still images is what is known as exposure bracketing, where multiple images of the same scene are captured sequentially, with varying exposure times, such that all parts of the scene are correctly represented in at least one exposure. One example of such a series of exposures is shown in Fig. 2.

This approach is often used for obtaining HDR photographs of even extreme scenes; however, as consecutive exposures are taken at slightly different times, fast-changing scenes (e.g., moving people or tree branches) can lead to misaligned exposures or ghosting artifacts, making it less suited to video. Nevertheless, certain video camera manufacturers [2] have employed this technique to a limited extent, optionally combining two exposures for extending the available dynamic range. To counter the temporal issues inherent with exposure sequences, differently exposed versions of the scene can be captured simultaneously using specialized cameras, where multiple sensors record the same scene at different exposure levels through optical systems that split light appropriately [3].

Despite these advances, the ideal solution for high quality video acquisition is using a single sensor. Typically, consumer cameras (such as the ones in smartphones) are limited by cost and physical size constraints, necessitating the use of small sensors that offer limited sensitivity. In contrast, professional cameras can incorporate larger sensors, allowing for an extended dynamic range to be captured with acceptable noise levels. Indeed, many professional cameras offer native HDR capture by making use of current HDR encoding formats (discussed in the following section). Nevertheless, similar capabilities are becoming available in certain digital SLR cameras in recent years (e.g., Sony alpha models offering native HDR video) or even in top-end smartphones (e.g., iPhone 12 Pro Dolby Vision video), democratizing HDR video acquisition.

Alternatively, if the HDR image itself is not required or cannot be displayed, multiple exposures can be directly combined into an SDR image, aiming to preserve as much of the original scene detail as possible while fitting within the more limited dynamic range of the SDR format. This process is known as *exposure fusion* [4] and

is commonly employed in smartphones under the provided “HDR” functionality to provide photographs with more detail, while avoiding to pass through a costly floating point format.

Encoding and Formats

An image or video encoding is defined by the precision it offers and the transfer function that it employs to translate linear luminance values to image intensities. Existing SDR encodings typically use 8 bits per color channel, providing 256 distinct levels. To distribute information in a more efficient manner, a nonlinear transfer function is used, translating scene values to image intensities. Largely motivated by historical reasons relating to the response of phosphors used in CRT television, the transfer function used for SDR imagery takes the form of an exponent with a value of 2.2–2.4.

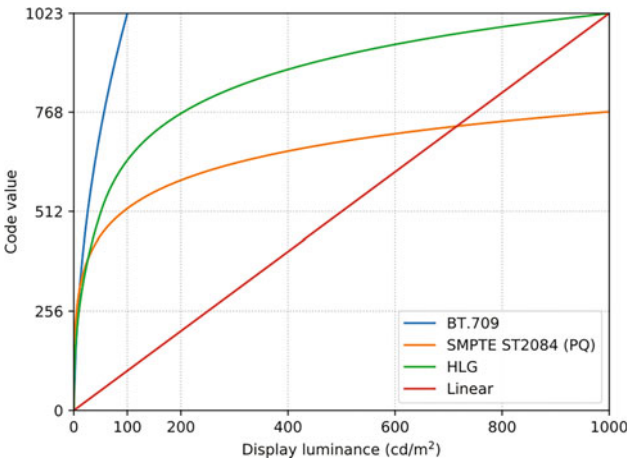
This encoding, standardized by the ITU [5], has remained prevalent until the development of HDR imaging and video but is not well-suited for content with a wider dynamic range, as consecutive intensity steps are likely to give rise to visible quantization artifacts. To alleviate this issue, HDR formats and encodings, as well as HDR-capable extensions to certain existing formats, have been developed.

Ideally, to represent HDR information in a lossless manner, a linear, floating point representation can be used, as is afforded by well-adopted OpenEXR format [6], with the consequence of

large file sizes. Nevertheless, EXR images are used both in HDR photography and as an intermediate format for film postproduction thanks to their near-lossless nature and the extension possibilities that the format offers. In an effort to reduce storage requirements, but without compromising quality, RGBE formats have been developed (standing for RGB and exponent), where 8 bits are used per color channel, as well as a shared exponent also encoded using 8 bits, which enables the conversion of 8-bit integer intensities to floating point with similar precision as the previously mentioned OpenEXR format. This encoding has been standardized in JPEG-HDR, where the RGB channels contain a tonemapped version of the image, offering the additional advantage of backwards compatibility [7].

In the case of video, although some efforts have been made in developing video formats for linear, floating point content, for example, through modification to MPEG video compression [8], to enable the use of HDR video content in use cases such as broadcasting or video streaming, more compact, integer-based formats are necessary to ensure compatibility with the more restricted hardware that is often found in consumer devices and bandwidth limitations. Several standards have been proposed to that end, defining how videos capturing a wider range of illumination can be encoded and subsequently displayed using integer-based encodings requiring 10 or 12 bits per color channel.

High Dynamic Range Imaging, Fig. 3 Transfer functions map between linear luminance values and digital code values



The most widely adopted formats are described in the ITU-R BT.2100 recommendation [9] and are the PQ format (standing for Perceptual Quantizer), developed by Dolby, and the HLG (Hybrid-log-gamma) encoding, developed by the BBC. Figure 3 shows a comparison of the encoding behavior of these formats relative to the BT.709 encoding used in SDR television and a linear encoding as might be used by the EXR format.

Display

The final stage within an HDR image pipeline is the display of still and video content, which can be achieved in two different ways. A number of display technologies have been developed for achieving extended ranges of illumination. The most common solution combines a spatially varying backlight with an LCD front panel, extending the dynamic range compared to conventional displays both toward the dark and the bright ends of the spectrum. Solutions exist that combine a DLP projector, placed behind an LCD panel [10] or alternatively, a matrix of LED light sources can replace the projector [11]. The latter solution has been adopted by television manufacturers in what is termed LED-backlighting and is commonly used in HDR-capable commercially available TVs. Finally, recent developments in display manufacturing have allowed of OLED displays to attain high enough luminance levels. This, combined with their ability to display true black as pixels can be individually turned off with minimal light leakage, makes them ideal candidates for HDR displays.

Although HDR display technologies are now commercially available, they are still only capable of displaying a fraction of the dynamic range available in nature. As such, to display images of arbitrary dynamic ranges on hardware with more restricted capabilities (including commercially available monitors), algorithms that can compress the range of the image to that of the display are necessary. Many such techniques exist and are known as *tonemapping operators*, which will be discussed in the following section.

Tone Reproduction

Different devices are capable of displaying varying dynamic ranges. To ensure that HDR images and content can be displayed on display devices with a more limited dynamic range, solutions allowing the HDR content to be mapped to the capabilities of a particular display are necessary. Since the output of such a mapping should still be useful for visual consumption, the form of this mapping is an important consideration.

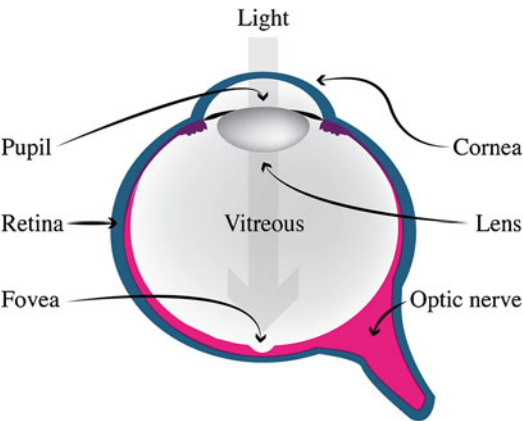
In the simplest scenario, one could linearly scale an HDR image to the given display range. Such a solution though would not be adequate for many scenes. Luminance values in HDR images are not linearly distributed across the available range. Instead, the distribution of luminance values in HDR imagery is highly kurtotic: sources and highlights only cover a small number of pixels in the image but are significantly brighter than the majority of the scene, while most pixels represent a relatively small range of intensities. In practice, the consequence of this is that a nonlinear mapping is necessary to ensure that sufficient detail in the scene remains visible.

The aim of tonemapping techniques can then be seen as the compression of the range of an HDR image such that it can be displayed on devices of a lower dynamic range, while preserving the visible detail in the scene. Often, this is also combined with quantization step, transitioning from linear floating point values to encoded, integer intensity values. Compressing the dynamic range of an image leads inevitably to the loss of some information. Consequently, tonemapping operators (TMOs) need to decide which information to keep and which to discard. The human visual system faces a similar challenge, as it is only capable of seeing a considerably smaller range than that available in nature. It should not come as a surprise then that dynamic range reduction is often inspired by aspects of human vision. Before discussing the different tonemapping techniques and their respective merit, a better understanding of the human eye

and the relevant processes to tone reproduction is necessary.

Perceptual Background

Human eyes are one of the means for sensing the world. Figure 4 shows an annotated cross section of the human eye. Light enters the eye through the pupil, travels through the ocular media, and finally hits the retina. The retina is a layer of neural sensors lining the back of the eye, which transduce light into a signal that is transmitted to the brain. Although several layers of neurons are present at the retina, it is the photoreceptors that are sensitive to light. Specifically, rod photoreceptors are sensitive to low light conditions, while cone photoreceptors are sensitive to brighter conditions.



High Dynamic Range Imaging, Fig. 4 A simplified cross section of the human eye

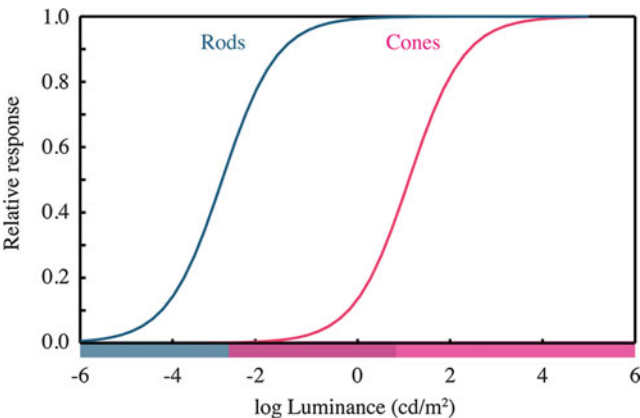
One of the most important properties of the human visual system with respect to dynamic range reduction is its ability to adapt to a wide range of illumination. This effect is experienced on a daily basis: it is sufficient to walk into a dark room from bright sunlight for the eyes to adjust to see under the new conditions. Effectively, the human visual system adapts to the prevailing lighting conditions so that contrasts in the scene can still be distinguished.

It is not a single process within the human visual system that is responsible for this effect, but rather a combination of factors. One familiar to most is the adjustment of the diameter of the pupil depending on the illumination. Another set of mechanisms occurring in the photoreceptors, which prevents them from saturating while adapting to the enormous range of illumination in nature, is their nonlinear response to illumination. This nonlinearity was first studied by Naka and Rushton [12] and provides a very effective mechanism for the compression of dynamic range. Intensities in the middle of the range lead to approximately logarithmic responses, while as the intensity is increased, the response tails off to a maximum. After that maximum is reached, further increases in intensity will not lead to a corresponding increase in response of the photoreceptors, ensuring that saturation will not occur [13].

Figure 5 shows the nonlinear response of photoreceptors to increasing luminance values, which can be modeled with what is known as the *Naka-Rushton equation*:

High Dynamic Range Imaging, Fig. 5

Photoreceptors respond nonlinearly to increasing luminance levels. The Naka-Rushton equation models this nonlinearity, producing S-shaped curves when plotted in a log-linear scale, known as sigmoid



$$\frac{V}{V_{\max}} = \frac{I^n}{I^n + \sigma(n)} \quad (1)$$

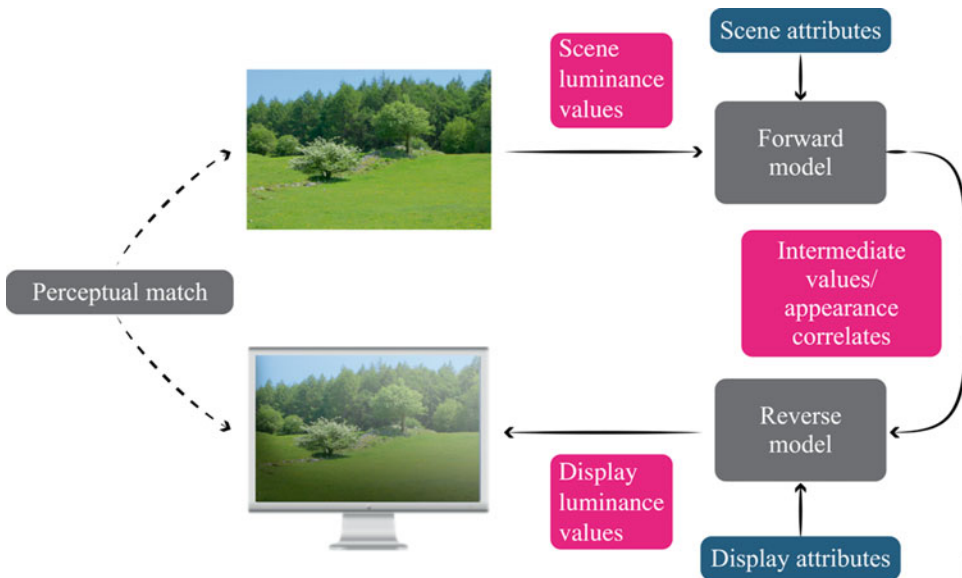
where V is the response at an intensity I , $V_{\max}$ is the peak response at saturation, and σ is the intensity necessary for the half-maximum response. The exponent n controls the slope of the function and is generally reported to be in the range of 0.7–1.0 [13]. This functional form is known as a *sigmoid*, because it forms an S-shaped curve when plotted on log-linear axes. This aspect of light adaptation of the visual system has played an instrumental role in dynamic range compression, which will be discussed in the following section.

Perceptually Motivated Tone Reproduction

Many tonemapping operators use models of photoreceptor adaptation to effectively reduce the dynamic range of input HDR images. Typically, this involves a forward step where the dynamic range of the image is compressed, considering the lighting conditions and appearance of the scene, followed by an inverse step that considers the parameters of the display where the image will be viewed, as shown in Fig. 6. If the forward step is derived from a model of the human visual system, it transforms the image from luminance

values to perceived values, essentially modeling what the original scene would be perceived as if the viewer were present. The reverse step then takes the compressed (now perceived) values back to luminance values, allowing them to be correctly displayed. Note however that many tonemapping operators do not employ the reverse step in order to maximize compression.

To approximate the compression afforded by photoreceptors, several functions have been proposed. At their simplest, linear scaling functions have been shown to be effective if the scaling factor is chosen appropriately, depending on the properties of the image and the intended display. Although such a linear compression scheme can create realistic results in less extreme cases, some information will be clamped for the image to fit within the displayable range. A better approximation of light adaptation behavior of photoreceptors can be achieved using a logarithmic mapping. Further, to better adapt this mapping to the image content, the base of the logarithms can be modified on a per-pixel basis in order to appropriately adjust the compression. More akin to photoreceptor behavior are sigmoid models, which have been used extensively in tone reproduction [14].



High Dynamic Range Imaging, Fig. 6 The typical steps in the tonemapping process

More complete models of photoreceptor physiology and light adaptation have been repeatedly used in the tone reproduction literature, offering perceptually accurate reproductions but at the cost of increased computational complexity and the need of accurate parameters describing the scene and viewing environment. An early and fairly complete model that incorporates luminance, color, and pattern representations to create a perceptually accurate compressed image employs a multi-scale processing scheme [15]. Another notable example that lends itself to HDR compression is the *retinex* model, which considers aspects of both the retina and the brain in order to model complex perceptual phenomena such as color constancy [16].

Multi-scale Tone Reproduction

Features in images occur at different scales, and many tone reproduction techniques employ mechanisms that treat images at multiple scales, improving the appearance of the final result as a different amount of compression can be applied to different features in the image. This ensures that high-contrast features are sufficiently compressed, while local edges (e.g., texture) are not proportionally scaled as that would reduce their visibility. To decompose the image into different scales, a form of edge-preserving filtering is typically used, such as the *bilateral filter* [17]. To improve upon the computational costs and limitations of the bilateral filter, other edge-preserving approaches have been proposed, including the domain transform [18] or the Laplacian pyramid [19].

Display and Viewing-Adaptive Tonemapping

Most tone reproduction algorithms discussed so far aim to compress the range of a given image to a displayable range, but with no consideration of any further properties of the device or the environment where the image will be viewed. Although many detailed appearance models exist that take into account aspects of the environment where the image was captured as well as where and how it will be displayed, such models tend to require additional input that is usually not available in a typical tonemapping setting.

In addition to the parameters of the viewing environment, different display devices offer varying capabilities that need to be considered. Mantiuk et al. formulated tonemapping as an optimization problem that strives to reduce the dynamic range of an image while minimizing the visible distortions that will inevitably be introduced in the process [20]. To detect visible distortions, a model of the human visual system is used, and to ensure that the resulting image is suitable for the target device, a display model is incorporated. Such a scheme allows for images to be tonemapped even for unconventional devices and viewing conditions.

Dynamic Range Expansion

Tonemapping techniques discussed so far compress the dynamic range of images to prepare them for displays with more limited capabilities. With the advent of HDR displays, the inverse problem arises. To fully take advantage of the extended dynamic range available on such a display, it may be desirable to expand the intensity range of traditional image and video content, a process known as *dynamic range expansion* or *inverse tonemapping*. This can be achieved by applying an expansion function to the intensity values of the image to map them to the range of the display. Both nonlinear (such the inverse of Eqs. 2 and 4) [21] and linear schemes have been proposed. Alternatively, an observation that has been exploited in inverse tonemapping is that in many scenes, luminance distribution is highly kurtotic, with most pixels covering a relatively small part of the range, while most dynamic range is allocated to light sources and highlights. To simulate this when expanding the dynamic range of images, bright areas of the image can be enhanced more [22].

To address more varied content, where the goal is not necessarily to recover the luminance distribution of the original real-world scene, several methods propose content-adaptive conversions, where the parameters used in the luminance expansion are determined by analyzing the characteristics of the content [23]. Focusing on the use

case of color-graded content, Bist et al. [24] proposed a style-aware adaptive solution, where the parameters are chosen so that the creative lighting style of the original is preserved as well as possible, as determined through a series of user studies.

A final issue to be considered in dynamic range expansion is that of under- or overexposed areas in the image. Even in a well-exposed image, light sources and highlights are likely clip to white, leading to loss of detail in those areas. When expanding such images, it is desirable to reintroduce some of the lost detail in these areas, which can be achieved through *hallucination* techniques, where overexposed regions are filled in using information from surrounding areas of the image [25]. More recently, the predictive power of deep learning approaches has been put to use for reconstructing the dynamic range of images in this context, focusing in particular on overexposed areas [26].

Commercial Solutions

Despite the extensive research in the field of dynamic range management, and the wealth of different approaches proposed in the literature, few commercial solutions are available for tackling the challenge of tone compression and expansion in professional video pipelines. Given the temporal constraints of real-time workflows (as would be the case in broadcasting, for instance) and the restricted hardware that might be encountered in embedded electronics (TVs, set-top boxes, cameras, etc.), complex, multi-scale filtering approaches are rarely implementable.

Further, the majority of existing solutions discussed above consider HDR imagery with a very wide dynamic range. In contrast, most media production workflows utilize HDR in a more conservative manner, with the goal of better preserving the highlights in the image, but without significantly increasing the average image luminance. Consequently, somewhat different approaches have been adopted in this kind of use case.

Several dynamic range mapping approaches are described in the ITU-R BT.2446 report [27]. Concretely, most existing luminance mappings used in production take the form of predefined look-up tables (LUTs) describing how the input

luminance information should be mapped, and consider specific reference levels. Alternatively, some content-adaptive solutions have been proposed commercially as discussed previously to ensure that the creative style of images is preserved despite changes in dynamic range.

Color Management in Tone Reproduction

Most of the tone reproduction operators discussed so far operate on a luminance representation of the image. Typically, a tonemapping operator will compress the world luminance L_w , which is computed from the RGB values in the image, to a display luminance L_d . To form the resulting tonemapped image, the RGB values will have to be scaled in such a way that the ratio between the three channels is kept the same before and after the compression. This can be achieved as follows:

$$\begin{bmatrix} R_d \\ G_d \\ B_d \end{bmatrix} = \begin{bmatrix} L_d \frac{R_w}{L_w} \\ L_d \frac{G_w}{L_w} \\ L_d \frac{B_w}{L_w} \end{bmatrix} \quad (2)$$

This process preserves the ratios between channels, but it is not sufficient to maintain the appearance of colors in the image. Saturation depends on both chromatic information and intensity values, and thus, luminance compression through tonemapping can lead to an oversaturated appearance (or undersaturated in the case of dynamic range expansion). To control the saturation of the tonemapped image, an exponent s can be used when reconstructing the three color channels from the adjusted luminance [28] (here shown only for the red channel):

$$R_d = L_d \left(\frac{R_w}{L_w} \right)^s \quad (3)$$

Although this approach is frequently employed as part of tonemapping operators, it may cause undesirable luminance shifts. Alternatively, the

three channels can be recomputed from the adjusted luminance as follows:

$$R_d = L_d \left(\left(\frac{R_w}{L_w} - 1.0 \right) s + 1.0 \right) \quad (4)$$

Both solutions offer simple saturation control but rely on manual selection for the parameter s . To automate the parameter selection for global tonemapping operators, a series of psychophysical experiments has derived a link between the gradient of tonemapping curve and the saturation parameter for both correction formulas [29]. An alternative, more general approach was proposed in Ref. [30], which can be applied without any knowledge of the tonemapping being necessary.

If additional information is available about both the image itself and the target viewing environment and device, accurate treatment of luminance in conjunction with color is possible using *color appearance models* (CAMs). Most such models are designed with accurate color reproduction as their primary aim and operate on a limited range of intensities. The increasing availability of HDR devices and content, however, has led to the development of several models capable of handling content of extended dynamic ranges, effectively combining tone reproduction and color appearance modeling. In contrast to the operators discussed so far, CAMs process each channel of the image separately rather than only compress luminance [31]. Typically, such models employ both a forward and an inverse step, taking the scene and viewing parameters into account, although it was recently shown that the inverse step may be bypassed, while still accurately reproducing color appearance [32].

Cross-References

- ▶ Adaptation
- ▶ CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams
- ▶ CIECAM02
- ▶ CIELAB
- ▶ Color Constancy

- ▶ Color Spaces
- ▶ Image Quality
- ▶ Psychological Color Space and Color Terms
- ▶ Unique Hues

References

1. Kunkel, T., Reinhard, E.: A reassessment of the simultaneous dynamic range of the human visual system. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization - APGV '10, p. 17 (2010). <https://doi.org/10.1145/1836248.1836251>
2. "RED HDRx." <https://www.red.com/red-101/hdrx-high-dynamic-range-video#w>
3. Tocci, M.D., et al.: A versatile HDR video production system. ACM Trans. Graph. **30**(4), 1–10 (2011). <https://doi.org/10.1145/2010324.1964936>
4. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: a simple and practical alternative to high dynamic range photography. Comput. Graph. Forum. **28**(1), 161–171 (2009). <https://doi.org/10.1111/j.1467-8659.2008.01171.x>
5. International Telecommunication Union. Parameter values for the HDTV standards for production and international programme exchange BT Series Broadcasting service. Recomm. ITU-R BT.709-5 **5**, 1–32 (2002) [Online]. http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.709-5-200204-I!!PDF-E.pdf
6. Kainz, F., Bogart, R., Hess, D.: The OpenEXR image file format. (2003)
7. Ward, G., Simmons, M.: JPEG-HDR: a backwards-compatible, high dynamic range extension to JPEG. In: SIGGRAPH 2006 – ACM SIGGRAPH 2006 Courses (2006). <https://doi.org/10.1145/1185657.1185685>
8. Mantiuk, R., Efremov, A., Myszkowski, K., Seidel, H.-P.: Backward compatible high dynamic range MPEG video compression. ACM Trans. Graph. (Proc. SIGGRAPH). **25**(3), 713–723 (2006)
9. International Telecommunication Union.: Image parameter values for high dynamic range television for use in production and international programme exchange. ITU-R BT.2100 (2016)
10. Seetzen, H., et al.: High dynamic range display systems. ACM Trans. Graph. **23**(3), 760 (2004). <https://doi.org/10.1145/1015706.1015797>
11. Seetzen, H., Whitehead, L.A., Ward, G.: 54.2: a high dynamic range display using low and high resolution modulators. SID Symp. Dig. Tech. Pap. **34**(1), 1450 (2003). <https://doi.org/10.1889/1.1832558>
12. Naka, K.I., Rushton, W.A.H.: The generation and spread of S-potentials in fish (Cyprinidae). J. Physiol. **192**(2), 437–461 (1967). <https://doi.org/10.1113/jphysiol.1967.sp008308>

13. Mayer, M.J., Dowling, J.E.: The retina: an approachable part of the brain. *Am. J. Psychol.* **101**(4) (1988). <https://doi.org/10.2307/1423238>
14. Reinhard, E., Ward, G., Pattanaik, S., Debevec, P.: *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*, 2nd edn. Elsevier Science (2010)
15. Pattanaik, S.N., Ferwerda, J.A., Fairchild, M.D., Greenberg, D.P.: A multiscale model of adaptation and spatial vision for realistic image display. In: *Proceedings of SIGGRAPH'98*, pp. 287–298 (1998). <https://doi.org/10.1145/280814.280922>
16. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**(6), 108–128 (1977). <https://doi.org/10.1038/scientificamerican1277-108>
17. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. (1998). <https://doi.org/10.1109/iccv.1998.710815>
18. Gastal, E.S.L., Oliveira, M.M.: Domain transform for edge-aware image and video processing. *ACM SIGGRAPH 2011 Pap. SIGGRAPH '11*. **1**(212), 1 (2011). <https://doi.org/10.1145/1964921.1964964>
19. Paris, S., Hasinoff, S.W., Kautz, J.: Local Laplacian filters: edge-aware image processing with a Laplacian pyramid. *ACM SIGGRAPH 2011*. **58**(3), 81–91 (2015). <https://doi.org/10.1145/2723694>
20. Mantiuk, R., Daly, S., Kerofsky, L.: Display Adaptive Tone Mapping. *ACM Trans. Graph.* **27**(3), 1 (2008). <https://doi.org/10.1145/1360612.1360667>
21. Banterle, F., Ledda, P., Debattista, K., Chalmers, A.: Inverse tone mapping. In: *Proc. Graph.* '06, p. 349 (2006). <https://doi.org/10.1145/1174429.1174489>
22. Didyk, P., Mantiuk, R., Hein, M., Seidel, H.P.: Enhancement of bright video features for HDR displays. *Comput. Graph. Forum.* **27**(4), 1265–1274 (2008). <https://doi.org/10.1111/j.1467-8659.2008.01265.x>
23. Kuo, P.H., Tang, C.S., Chien, S.Y.: Content-adaptive inverse tone mapping. (2012). <https://doi.org/10.1109/VCIP.2012.6410798>
24. Bist, C., Cozot, R., Madec, G., Ducloux, X.: Tone expansion using lighting style aesthetics. *Comput. Graph.* **62**, 77–86 (2017). <https://doi.org/10.1016/j.cag.2016.12.006>
25. Wang, L., Wei, L.-Y., Zhou, K., Guo, B., Shum, H.-Y.: High dynamic range image hallucination. In: *Eurographics Symposium on Rendering 2007*, pp. 321–326 (2007). <https://doi.org/10.2312/EGWR/EGSR07/321-326>
26. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **36**(6) (2017). <https://doi.org/10.1145/3130800.3130816>
27. International Telecommunication Union-Radio-communication (ITU-R): Report BT.2446-0: methods for conversion of high dynamic range content to standard dynamic range content and vice-versa, vol. 0, 2019
28. Tumblin, J., Hodgins, J.K., Guenter, B.K.: Two methods for display of high contrast images. *ACM Trans. Graph.* **18**(1), 56–94 (1999). <https://doi.org/10.1145/300776.300783>
29. Mantiuk, R., Mantiuk, R., Tomaszewska, A., Heidrich, W.: Color correction for tone mapping. *Comput. Graph. Forum.* **28**(2), 193–202 (2009). <https://doi.org/10.1111/j.1467-8659.2009.01358.x>
30. Pouli, T., Artusi, A., Banterle, F., Akyüz, A.O., Seidel, H.-P., Reinhard, E.: Color correction for tone reproduction. In: *Final Progr. Proc. - IS T/SID Color Imaging Conf.*, pp. 215–220 (2013)
31. Fairchild, M.D.: *Color Appearance Models*. Wiley (2013)
32. Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A., Damberg, G.: Calibrated image appearance reproduction. *ACM Trans. Graph.* **31**, 1 (2012). <https://doi.org/10.1145/2366145.2366220>

High Pressure Mercury Lamp

Wout van Bommel

van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

High-pressure mercury gas-discharge lamp;
High-pressure mercury-vapor lamp

Definition

A lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a high-pressure mercury vapor that is contained in a transparent bulb. In some versions, fluorescent powder is applied that converts the ultraviolet part, which is emitted together with visible light, into visible light to improve the color quality of the light.

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

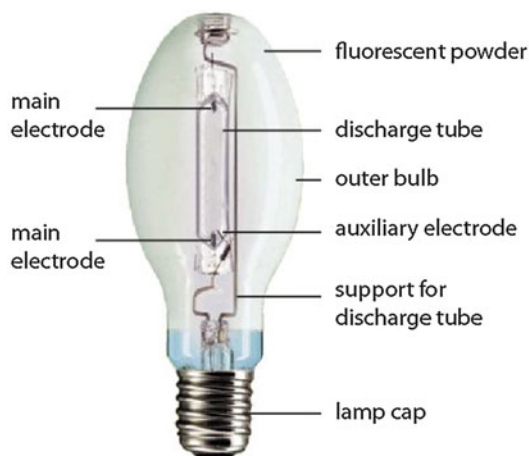
High-Pressure Mercury Gas-Discharge Lamps

High-pressure mercury lamps belong to the group of high-intensity discharge (HID) lamps because they are available in high lumen output (and thus high luminous intensity) versions. High-pressure mercury lamps are available in versions where the discharge takes place in vaporized mercury only and in versions in which metal halides are added so that the discharge takes place in mercury vapor and in vaporized metals from the metal-halide components [2, 3]. These latter types are called metal-halide lamps. Given the unique operating principle and construction of metal-halide lamps, they are dealt with in a separate entry “metal-halide lamp.”

High-pressure mercury lamps, like all high-pressure discharge lamps, are compact compared to low-pressure discharge lamps. They have moderate efficacy and moderate color rendering. With their cool-white light, they were, until the 1970s of the last century, extensively used in road lighting, especially in built-up areas. Since the introduction of the more efficient high-pressure sodium lamps in the late 1960s, these lamps have in many cases replaced high-pressure mercury lamps.

Working Principle

The gas-discharge principle of high-pressure mercury lamps is similar to that of all other gas-discharge lamps. In high-pressure mercury lamps, the discharge takes place in vaporized mercury at a pressure of around 10^6 Pa (10 atm). The spectrum of the radiation is a line spectrum with emissions in the long-wave UV region and in the visible region at the yellow, green, blue, and violet wavelengths. The lamp without fluorescent powder lacks red in its spectrum and has a bluish-white color appearance and very poor color rendering. In most high-pressure mercury lamps, fluorescent powders are used to improve the color quality by converting a large part of the (small) UV component into visible radiation, predominantly in the red end of the spectrum. The result is cool-white light of moderate color rendering and improved efficacy. Like all gas-



High Pressure Mercury Lamp, Fig. 1 Principle parts of a high-pressure mercury gas-discharge lamp [1]

H

discharge lamps (with very few exceptions), a high-pressure mercury lamp cannot be operated without a ballast to limit the current flowing through it.

Materials and Construction

The main parts of a high-pressure mercury lamp are (Fig. 1):

- Discharge tube
- Fill gas
- Electrodes
- Outer bulb (often with fluorescent coating)
- Lamp cap

Discharge Tube

In view of the high operating temperature, quartz is used for the discharge tube because it has a higher melting temperature than glass.

Fill Gas

The discharge tube contains a small quantity of mercury and an inert gas filling.

Electrodes

The main electrodes consist of a core of tungsten rod with a tungsten coil (impregnated with emissive material) wound around it. To aid starting, an ordinary high-pressure mercury lamp has not only an inert gas but also an auxiliary electrode.

Because of this, an ordinary mercury lamp does not need an external igniter. The auxiliary electrode simply consists of a tungsten wire positioned very close to one of the main electrodes (Fig. 2).

Outer Bulb

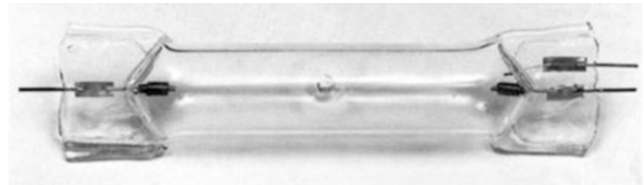
An outer bulb (usually ovoid in shape) with an inert gas filling isolates the gas-discharge tube so that changes in ambient temperature have no influence on its proper functioning. It also protects the lamp components from corrosion at the high operating temperatures involved. For the smaller wattage lamps, with their lower operating temperatures, ordinary glass is used, while for the other types, hard glass is used.

Fluorescent Powder

As has already been mentioned, high-pressure mercury lamps usually employ fluorescent powder to improve the color quality of the light emitted. The powder is provided as a coating on the inner surface of the outer bulb. Different fluorescent coatings are used to obtain different lamp types with different color qualities and lamp efficacies.

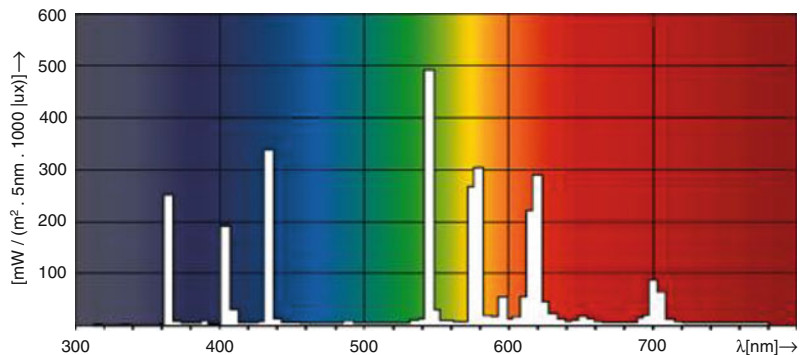
High Pressure Mercury Lamp,

Fig. 2 Discharge tube showing the auxiliary electrode and two main electrodes



High Pressure Mercury Lamp,

Fig. 3 Spectral energy distribution of a high-pressure mercury lamp with fluorescent powder [1]



Lamp Cap

Lamp caps are of the Edison screw type, with the wattage of the lamp determining their size (E27 and E40).

Properties

Energy Balance

Approximately 17% of the input power is emitted in the form of visible radiation. Compare this with the 28% of a tubular fluorescent lamp and the 30% of a high-pressure sodium lamp.

Luminous Efficacy

Luminous efficacy varies with lamp wattage and with the color quality of the lamp from some 35–60 lm/W.

Lumen Package Range

High-pressure mercury lamps are produced in lumen packages between some 2,000 and 60,000 lm (corresponding wattages between 50 and 1,000 W).

Color Characteristics

As has already been mentioned, high-pressure mercury lamps have a line spectrum (Fig. 3).

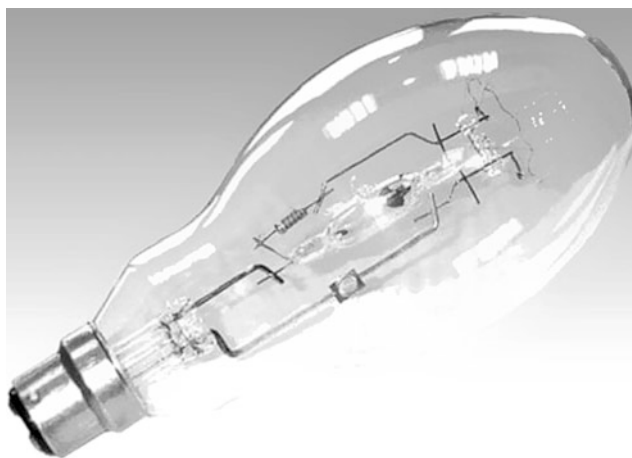
High Pressure Mercury Lamp,

Fig. 4 Different wattage ovoid and reflector type of high-pressure mercury lamps [1]



High Pressure Mercury Lamp,

Fig. 5 Blended light lamp with high-pressure mercury gas-discharge tube and tungsten incandescent filament combined in one outer bulb. To show the construction, a clear lamp is shown, although most blended light lamps use a fluorescent powder coating on the interior of the outer bulb [1]



H

The two lines in the red part of the spectrum are obtained by conversion of ultraviolet radiation by the fluorescent powder. The color characteristics are dependent on the composition and quality of the fluorescent powders used. Different compositions and qualities are used to produce lamps with color temperatures between some 3,500 and 4,500 K, with color rendering index (R_a) values of around 60 for high-quality versions and around 40 for ordinary versions.

Lamp Life

As with most gas-discharge lamps, lamp life is determined by emitter exhaustion of the electrodes. Economic life varies according to the type between 10,000 and 15,000 h (20% mortality).

Lamp Lumen Depreciation

Lamp lumen depreciation is caused by evaporation and scattering of electrode material (lamp

blackening) and by the gradual decrease in the activity of the fluorescent powder. The point at which 20% lumen depreciation occurs lies at around 10,000–15,000 h.

Run-Up and Reignition

The run-up time of a high-pressure mercury lamp to its full temperature and corresponding nominal mercury pressure is some 4 min. The hot lamp will not restart until it has cooled sufficiently to lower the vapor pressure to the point at which restrike with the voltage available is possible. The reignition time is in the order of 5 min.

Dimming

High-pressure mercury lamps cannot be dimmed.

Mains Voltage Variations

A 5% variation in the mains voltage changes both lamp current and light output by 10%.

Overvoltage decreases the lamp life and increases lamp depreciation because of the correspondingly higher current.

Product Range

High-pressure mercury lamps are available in an ordinary version with poor color rendering (R_a of around 40) and in so-called comfort versions with an improved color rendering of around 60. The bulb is ovoid in shape and increases in size, with an increase in wattage (Fig. 4). Versions without fluorescent powder are hardly, anymore, produced. Reflector lamp versions are produced with a cone-shaped outer bulb and an internally reflective coating on the front (Fig. 4, right).

There is one version of the high-pressure mercury lamp, the “blended light lamp,” that does not need an external ballast. The ballast has simply been built into the lamp itself in the form of a tungsten filament (Fig. 5). The lamp can be connected directly to the mains. The light from the mercury discharge and that from the heated filament blend together (hence the name blended light lamp). The color characteristics of this lamp are therefore better than those of an ordinary high-pressure mercury lamp, but this comes at the cost of a considerably lower efficacy.

Cross-References

- [Metal Halide Lamp](#)
- [Phosphors and Fluorescent Powders](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Coursebook. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. Illuminating Engineering Society of North America (IESNA), New York (2011)

Highly Automated Vehicles

- [Autonomous Vehicle \(AV\) Lighting Systems](#)

High-Pressure Mercury Gas-Discharge Lamp

- [High Pressure Mercury Lamp](#)

High-Pressure Mercury-Vapor Lamp

- [High Pressure Mercury Lamp](#)

Highway Lighting

- [Road Lighting](#)

History, Use and Performance

- [CIELAB](#)

Hue Difference, Delta H

- [CIE94, History, Use, and Performance](#)

Hue Opponency

- [Color-Opponency, Unique Hues](#)

Hue Sequence

- [Color Circle](#)

Human-Centric Lighting

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

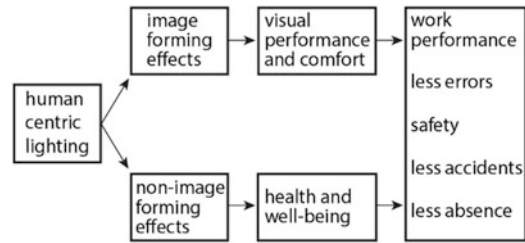
[Integrative lighting](#)

Definition

Human-centric lighting (HCL Lighting) that takes both the beneficial visual and nonvisual biological effects of lighting into account. CIE defined in a draft international standard the term “Integrative Lighting” with the same meaning as human-centric lighting: “lighting specifically designed to produce a beneficial physiological and/or psychological effect upon humans, including visual and non-visual effects” [1].

Image and Non-Image Forming Processes in the Eye

An image forming process in the eye produces visual effects of light. Other processes produce nonvisual biological effects of light. They are, therefore, sometimes referred to as “non-image forming effects of light” (in short, NIF effects of light). Today, most lighting standards in the world are developed so that the lighting installation provides proper visual effects seen from both a point of view of visual performance and visual satisfaction (visual comfort). The non-image forming effects influence the way our body “operates” and therefore, influence our health and well-being (see the entries ► [“Nonvisual Lighting Effects and Their Impact on Health and Well-Being,”](#) ► [“Circadian Rhythms,”](#) and ► [“Sleep, Alertness, and Light”](#) of this Encyclopedia). Only if the lighting designer succeeds in designing and installing lighting that results in both suitable visual and nonvisual biological effects,



Human-Centric Lighting, Fig. 1 Visual and nonvisual biological effects of lighting and benefits of good lighting

optimised visual performance and comfort, health and well-being may be expected to be obtained. In the context of lighting for work, this, in turn, propagates better work performance, fewer errors, improved safety, and lower absenteeism. Figure 1 sketches the relationship between the different effects of lighting and shows the benefits to be expected from good lighting for work. Good human-centric lighting can have a positive influence on work performance, errors, safety, accidents, and absenteeism.

Circadian and Direct Photobiological Effects of Light

Non-image forming effects of light consist of circadian and direct photobiological effects. Daily or circadian bodily rhythms, a fundamental property of human life, are synchronized by the natural 24-h dark-light rhythm. Light and darkness have a long-term effect on circadian bodily rhythms [2–6]. A correct circadian rhythm ensures, among others, that melatonin, the hormone that increases sleepiness, has its maximum concentration slightly after midnight, while cortisol, the hormone providing energy to the body, has its maximum in the morning. Light received up to a few days before can influence the circadian rhythm. As such, it has a long-term effect. Light can, however, also have short-term, direct effects within a short time after being exposed to light. The effects usually disappear shortly after the light exposure stops. Light, both at day and night, can acutely decrease sleepiness and improve alertness [7–9]. The increase of alertness, in turn, has positive consequences for

performance. Both effects of light are discussed in more detail in the entries ► [“Nonvisual Lighting Effects and Their Impact on Health and Well-Being,”](#) ► [“Circadian Rhythms,”](#) and ► [“Sleep, Alertness, and Light”](#) of this Encyclopedia.

Dynamic Lighting Scenario for Daytime Work

The state of knowledge of today indicates that for entrainment of the biological clock and for direct photobiological effects (influencing alertness and performance) the bare minimum illuminance at the eye is some 400 to 500 lux. (see the entries ► [“Nonvisual Lighting Effects and Their Impact on Health and Well-Being,”](#) ► [“Circadian Rhythms,”](#) and ► [“Sleep, Alertness, and Light”](#) of this Encyclopedia). This holds for light sources with a correlated color temperature between approximately 3000 and 4000 K. For higher color temperatures, the required value is lower. For values of around 6500 K, the required value can be halved: some 200 to 250 lux. Lamps of 3000 to 4000 K have a melanopic equivalent daylight (D65) factor of some 0.4 to 0.6 and lamps of 6500 K approximately 1.0. These eye illuminances, therefore, correspond to melanopic equivalent daylight (D65) illuminances of approximately 200 to 300 lux.

The nonvisual biological effect of light is not directly governed by the illuminance on the working plane, but by light entering the eye. For the lighting designer, however, it is more convenient to work with illuminances on the horizontal working plane. Converting required illuminances on the eye (for all locations in the space) into an approximate average horizontal illuminance value of the horizontal working plane is thus needed. With the kind of light distributions and layout of luminaires typically used in office and industrial spaces, the average horizontal illuminance at working plane level corresponds to roughly three times the amount of vertical illuminance at eye height. On this basis, Table 1 gives the minimum required average

Human-Centric Lighting, Table 1 Minimum horizontal illuminance on the working plane of office type of spaces. Converted from illuminances at the eye

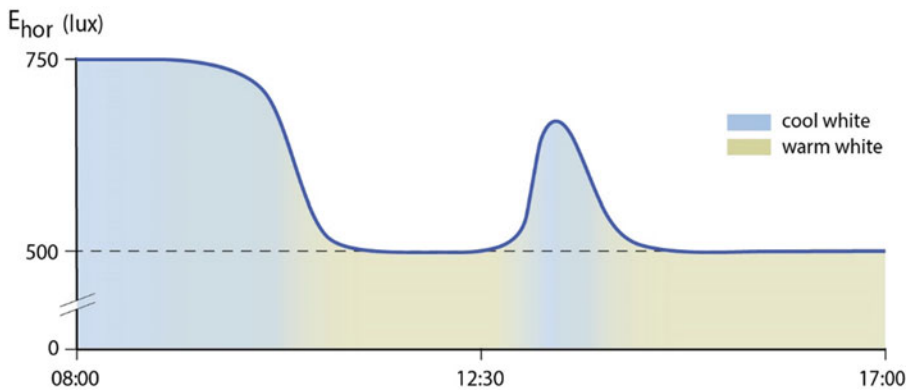
CCT	E _{eye} (lux)	E _{eye, mel, D65eq} (lux)	E _{hor, av} (lux)
3000–4000 K	500	250	1500
ca. 6500 K	250	250	750

horizontal illuminance values on the working plane.

In most offices and industrial spaces, these lighting levels are not available from daylight entering the premises for those not working close to windows, especially not in wintertime outside the tropical latitudes. In these situations, artificial lighting should take care of the extra lighting required. A lighting scenario which dynamically changes both the lighting level and color enables optimization between energy requirements on the one hand and requirements of visual and nonvisual effects on the other hand.

Figure 2 sketches an example of such a daytime dynamic lighting scenario for a working day starting at 8:00 in the morning and finishing at 17:00.

The morning lighting begins with a horizontal illuminance level of 750 lux and a cool-white color (approx. 6500 K). This situation corresponds to roughly 250 lux melanopic equivalent daylight (D65) illuminance. It helps to entrain the biological clock and evokes direct photobiological effects providing alertness and performance potential. This lighting condition lasts for some 2.5–3 h. After that, the lighting level gradually decreases, and the color gradually turns warmer. By the beginning of the afternoon, around lunchtime, the lighting level reaches its lowest level of 500 lux while the color has its warmest tint (around 3000 to 4000 K). This situation corresponds to roughly 85 lux melanopic equivalent (D65) illuminance. The minimum horizontal lighting level is never lower than 500 lux, being the minimum level recommended for visual performance and comfort. The midday situation of low lighting level and warm color supports relaxation around lunchtime. It also does not work



Human-Centric Lighting, Fig. 2 Example of a dynamic lighting scenario for human-centric lighting in offices (daytime use)

against short naps for those who want to have them. Lunchtime power naps have shown to help to feel revitalized and to improve performance for some 2–3 h in the afternoon [10, 11]. These multiple short naps should last some 3–20 min and be taken well before 15:00 h. The latter, in order not to diminish the sleep pressure, is already building up and required for sleep during the night. A short bright light exposure after lunchtime also has a revitalizing effect [12, 13]. Therefore, a short-lasting (some 30 min) post-lunch rise in light level combined with higher color temperature is included in the scenario of Fig. 2. After the post-lunch rise, the lighting level decreases again while the color temperature gradually returns to 3000 to 4000 K. The absence of bright cool light in the afternoon also avoids interference with the gradual rise of melatonin later in the afternoon or early evening.

The dynamic daytime scenario illustrated in Fig. 2 is based on the bare minimum requirements for everyday office tasks. For difficult visual tasks, the minimum lighting level has to be higher than 500 lux. For challenging tasks with severe consequences in the case of errors, alertness and performance requirements can be extremely high. This, of course, also holds for potentially hazardous conditions in industrial environments. For such tasks or conditions, the lighting level of 750 lux should be increased, and higher correlated color temperatures than 6500 K could be considered.

Cross-References

- [Circadian Rhythms](#)
- [Melanopic Equivalent Daylight Illuminance](#)
- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)
- [Sleep, Alertness, and Light](#)

References

1. CIE (2016) CIE DIS 017:2016 (term 17-29-030). CIE Draft international standard, ILV: International lighting vocabulary, CIE, Vienna
2. Chovnick, A.: Proceedings of Cold Spring Harbor Symposium on Quantitative Biology. Biological Clocks 25 (1960)
3. Aschoff, J.: Circadian rhythms in man – a self-sustained oscillator with an inherent frequency underlies human 24-hour periodicity. *Science*. **148**, 1427–1432 (1965)
4. Boivin, D.B., Czeisler, C.A.: Resetting of circadian melatonin and cortisol rhythms in humans by ordinary room light. *Neuroreport*. **9**, 779–782 (1998)
5. Foster, R., Kreitzman, L.: Rhythms of Life. The Biological Clocks that Control the Daily Lives of every Living Thing. Profile Books Ltd., London (2004)
6. Roenneberg, T., Juda, M., Kramer, A., Mellow, M., Vetter, C., Allebrandt, K.V.: Light and the human circadian clock. In: Kramer, A., Mellow, M. (eds.) *Circadian Clocks*. Springer, Berlin Heidelberg (2013)
7. Vandewalle, G., Balteau, E., Phillips, C., Degueldre, C., Moreau, V., Sterpenich, V., Albouy, G., Darsaud, A., Desseilles, M., Dang-Vu, T.T., Peigneux, P., Luxen, A., Dijk, D.-J., Maquet, P.: Daytime light exposure dynamically enhances brain responses. *Curr. Biol*. **16**, 1616–1621 (2006)

8. Lockley, S.W., Evans, E.E., Scheer, F.A.J.L., Brainard, G.C., Czeisler, C.A., et al.: Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep*. **29**, 161–168 (2006)
9. Cajochen, C.: Alerting effects of light. *Sleep Med. Rev.* **11**, 453–464 (2007)
10. Mednick, S.C., Ehrman, M.: *Take a Nap! Change Your Life*. Workman Publishing Company, New York (2006)
11. Ficca, G., Axelsson, J., Mollicone, D.J., Muto, V., Vitiell, M.V.: Naps, cognition and performance. *Sleep Med. Rev.* **14**, 249–158 (2010)
12. Kaida, K., Takahashi, M., Haratani, T., Otsuka, Y., Fukasawa, K., Nakata, A.: Indoor exposure to natural bright light prevents afternoon sleepiness. *Sleep*. **29**, 462–469 (2006)
13. Kaida, K., Takeda, Y., Tsuzuki, K.: The effects of short afternoon nap and bright light on task switching performance and error-related negativity. *Sleep Biol. Rhythms*. **11**, 125–134 (2013)

Hyperspectral Imaging

Juan Luis Nieves

Optics Department, Science Faculty, University of Granada, Granada, Spain

Synonyms

[Color imaging](#); [Color science](#); [Multispectral imaging](#); [Spectral imaging](#)

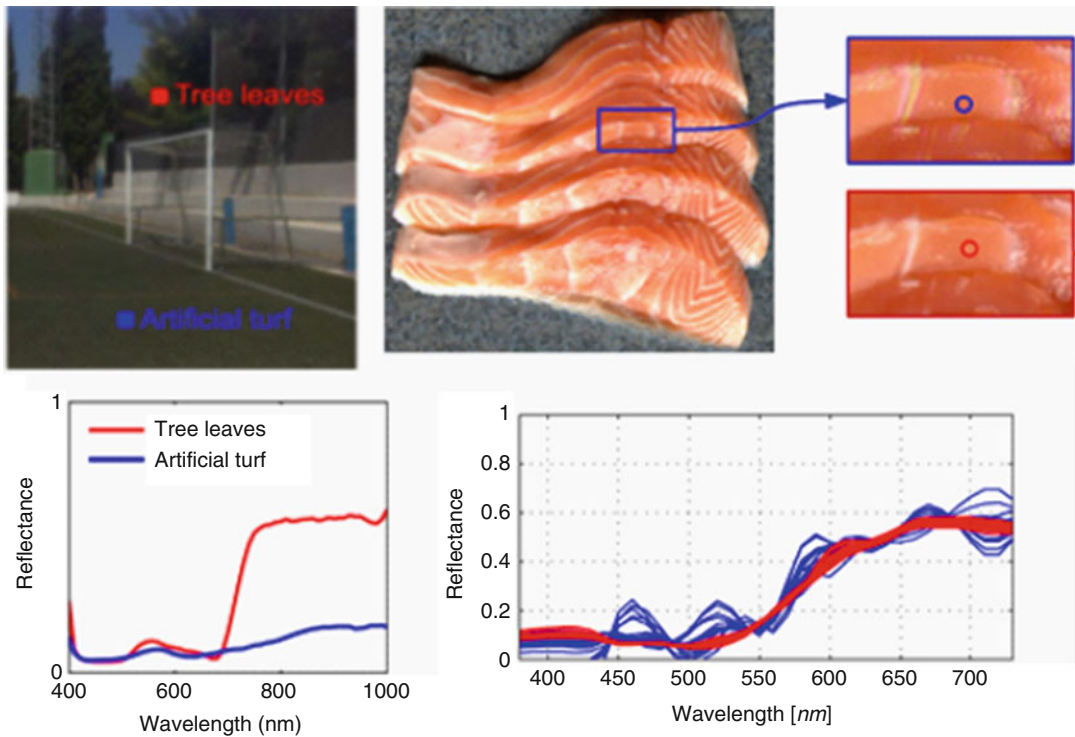
Definition

The goal of spectral imaging is to recover radiance and reflectance spectra at each pixel in a scene. The stored hyperspectral information for each pixel refers to data from spectra which is sufficiently well scanned so that to be considered as a continuous spectral signal. Typically, such spectral imaging systems consist of a digital camera coupled to a range of spectrally broadband or narrowband filters. Depending on the number of filters used, spectral systems can be classified into multispectral imaging systems, when a small number of 3–10 filters is used, hyperspectral systems, if the number of filters extends up to

10–100, or ultraspectral, when the number is even greater than 100. If the number of filters is sufficiently large and their bandwidths are sufficiently small, as with a hyperspectral and ultraspectral imaging systems [1–3], spectral data can be recovered exactly.

Overview

Multispectral imaging ordinarily uses a digital charge-coupled device (CCD) camera coupled with colored filters of different spectral bands, ranging from just three components, as in a red-green-blue (RGB) conventional camera, up to hundreds of components, as in a hyperspectral system [3]. If the sensor of the CCD camera is monochrome, a high number of narrowband filters is traditionally used, while if the sensor is the RGB color type broadband filters can be used which makes it possible to work with a significantly reduced number of filters. It is the type of sensor which determines the number of channels in a spectral camera thus allowing a first classification of the different typologies of the acquired spectral images, whether with multi- or hyperspectral imaging. The main application of spectral imaging is in areas where a good or high spatial resolution of the spectral measurement is needed. Traditionally, conventional spectroradiometers are the devices used for spectral measurements which are not only expensive but just allow a spectral measurement of a relatively large area (some 2°–10°) of the scene. This significantly limits the spatial details of the scene to be spectrally characterized. On the other hand, the spectral devices used for acquiring images allow, for example, the measurement of the spectral reflectance of each pixel in the scene (Fig. 1). The hyperspectral techniques used for acquiring images are very important in color imaging because they allow a full spectral reflectance or spectral radiance measurement for each pixel image, thus “preserving” almost the exact colorimetric components of the scene, by just converting this spectral radiance or reflectance to the CIE XYZ tristimulus values [4, 5].



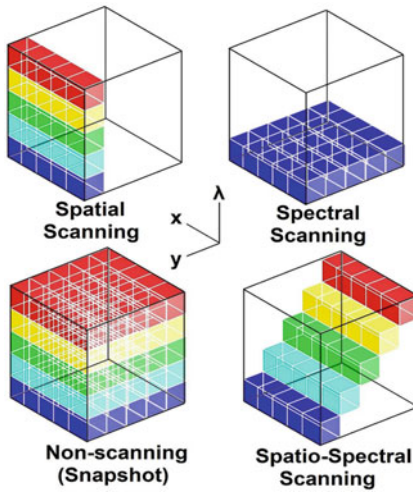
Hyperspectral Imaging, Fig. 1 Rendered scenes (top row) and (bottom row) captured spectra from a commercial Bragg-grating-based hyperspectral imager and a

commercial multispectral camera truePIXA (<https://www.chromasens.de/en>)

Some Basics on Image Acquisition and Spectral Imaging

Image capture devices are used in general to obtain digitized images of real-world scenes by forming a real image of the scene into its sensor plane area. Thus, their purpose is to transform the continuous, wide range, and very detailed real world into a set of flat quantized values, each corresponding to a pixel. For a conventional RGB color image, this process is equivalent to building a stack of three planes (or sets of organized pixels), each one corresponding to the red, green, and blue dimensions. The camera sensor can either be a CCD (charged-coupled device, known as the technology of moving and storing the electron charge) or a CMOS (complementary metal oxide semiconductor, known as the technology used to build a transistor into a semiconductor wafer). Typically, an organized array of small square color filters called a Bayer filter is placed in front of the sensor to allow

color image capture. The raw data from the sensor are fed into the image processor which performs the transformation of the raw data to the quantized three-plane color image. For most color digital cameras, the color is read by placing a mask of colored filters in front of the sensor, and then interpolating the readings of adjacent pixels to get the three RGB values. The fact is that standard color digital cameras never have full-color resolution, as each pixel is obtained by interpolating from adjacent pixels, and this is the cause of some artifacts such as color fringing in the image in some cases. The interpolation procedure also makes the capture process slower. Moreover, the usual practice in color imaging is to modify the natural spectral response of the sensor by placing an infrared-blocking filter to cut-off the higher wavelength response, and also of course the three RGB filters (either in a Bayer arrangement or covering the whole area of the sensor). In the end, the color cameras have three spectral response curves, each



Hyperspectral Imaging, Fig. 2 Illustration about how to store the spatial and spectral information in different hyperspectral imaging systems

corresponding to a red, green, or blue color channel.

A hyperspectral image of n bands (or n channels) with spatial resolution of $(r \times c)$ can be represented as a three-dimensional data cube $\mathbf{C} = [C(x; y; \lambda)]_{(r \times c \times n)}$, where x and y represent the spatial dimensions of the images, and λ represents the spectral dimension; in most cases, the image intensity corresponds to either the spectral radiance or spectral reflectance factor [3, 5, 6]. As can be seen in the next sections, there are many ways of capturing that spectral cube. Some spectral imaging systems need dispersive optics elements (a prism or a single slit) to spectrally decompose the incoming light. Alternatively, other system architectures use a set of a number of optical filters mounted in a filter wheel placed sequentially in front of a monochrome camera. For the purpose of a proper calibration, raw image sequences have to be corrected for noise (e.g., camera dark-current noise and non-uniformity in the sensor array response), stray light, off-axis vignetting, and potential chromatic differences of magnification or translation. The spectral reflectance at each pixel can be estimated by normalizing the aforementioned corrected signals against that obtained from a neutral reference flat-field placed in the captured scene [3, 5]. As

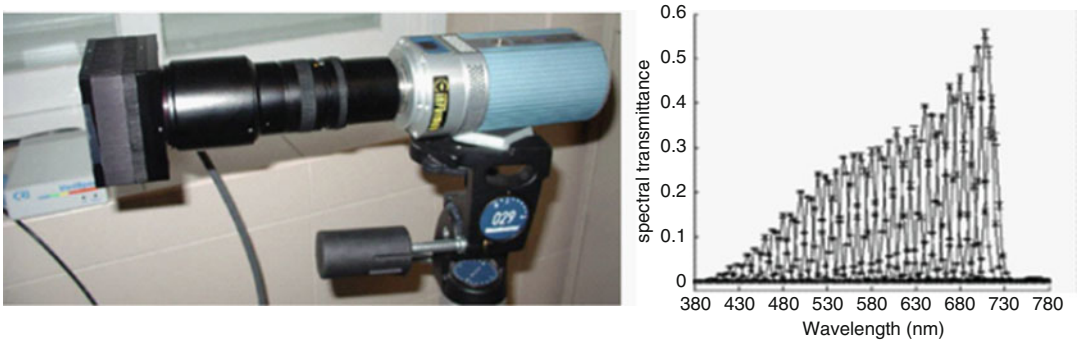
illustrated in Fig. 2, the spectral data collected are stored as a “hypercube” $\mathbf{C}$, which contains spatial data in two dimensions and spectral data along a third dimension for each pixel in the image.

Direct Spectral Measuring Systems

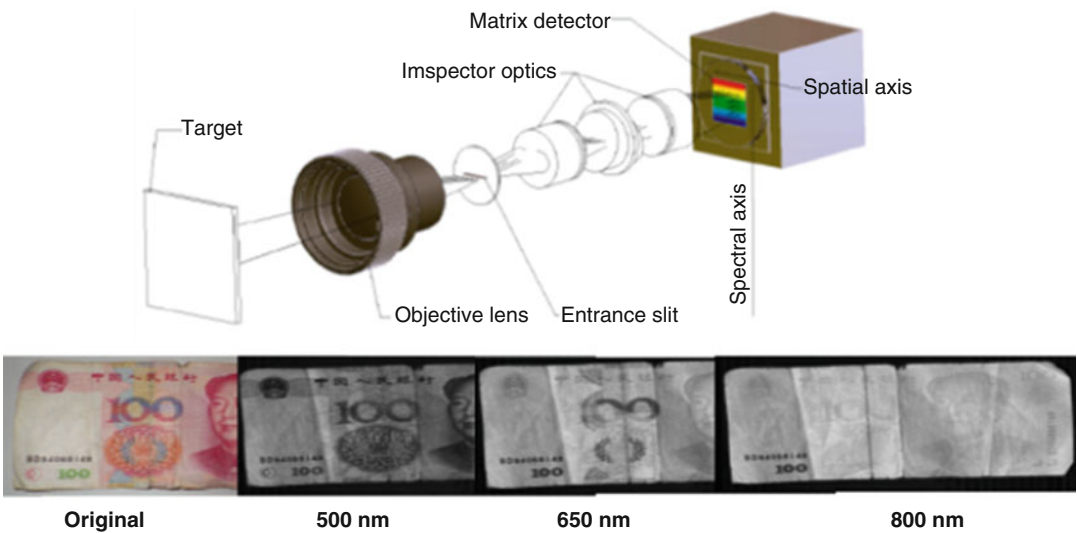
These systems are conventionally known as hyperspectral imaging systems. If the number of channels n is sufficiently large and their bandwidths are sufficiently small, spectral data can be recovered exactly from the data cube $\mathbf{C}$. Thus, these systems require both a huge amount of processing time and a large storage capacity [7]. Such hyperspectral systems in turn can be classified in three main categories, depending on the scanning step introduced along the spectral and spatial dimension: spectral domain scanning systems, spatial domain scanning systems, and snapshot-based systems.

Spectral Domain Systems

These spectral systems scan an image in the spectral domain by separating the light into its spectral components. To this end, they use a monochrome sensor coupled to a set of narrowband filters. To record a complete hyperspectral cube $\mathbf{C}$, several image captures of the scene are needed (one per spectral band), which means that the objects within the scene should remain static. One of the most used narrowband filters used in these types of systems is the liquid crystal tunable filter (LCTF). The LCTF is a tunable interference filter based on the electro-optic effect present in liquid crystals (or crystals). In short, it consists of two polarizers and a variable retarder placed between them, which changes its refraction index when an electrical field (voltage difference) is applied which provokes an electro-optic effect. The LCTF illustrated in Fig. 3 was manufactured by the Varispec series of CRI with a usable wavelength range from 400 to 720 nm, and the optical resolution or bandwidth of the spectral transmittance peaks was from 7 to 20 nm. This device allowed a tuning resolution of 1 nm and thus is equivalent to the use of a large set of narrowband filters. In the case of outdoor hyperspectral



Hyperspectral Imaging, Fig. 3 (Left) CCD monochrome camera coupled with an LCTF placed in front of the lens of the camera; (right) spectral sensitivities of the channels selected in the LCTF [9]



Hyperspectral Imaging, Fig. 4 Working principle of the Specim's Inspector (top row) and (bottom row) example of a scene captured in three spectral bands (from <https://www.specim.fi/>)

imaging, varying daylight illumination has to be taken into account [8, 9].

Spatial Domain Systems

These systems are based on a local scan of the areas of the scene in order to obtain the complete spectral information of each of these areas. The most common configuration to achieve this scan is the use of line-scan CCD sensors which allow a line by line sweep of the scene, resulting in a high spatial resolution. A line-scan camera scans moving objects by repeatedly capturing a single line of pixels at a high frequency. Therefore, these spectral systems act like line scanners which disperse

the polychromatic light emerging from each pixel and separately register each wave length component of each point in the scanned line. One of the most commonly used setups is Specim's Inspector (see Fig. 4) where either the sample or the camera has to be moved to achieve the special scan required. This device is commonly used in the quality inspection industry to monitor the quality of fruit and foodstuffs and for the accurate color classification of the products.

Snapshot-Based Systems

These systems are able to spectrally decompose the incoming light from very small areas of the

scene. Thus, in the sensor array, they receive both bidimensional spatial and spectral information. This is a clear advantage in comparison with other spectral systems but due to the complex pupil and lens arrays used to reimage the dispersed areas of the scene in the sensor, the sensor resolution is poor (around 300×300 pixels) and needs a powerful controlled illumination. These systems are thus mainly used in microscopy [10].

Indirect Spectral Imaging Systems

When a CCD digital color camera is pointed at a surface with a spectral reflectance function $r(x, y; \lambda)$, the response of the n th sensor for pixel (x, y) can be modeled linearly by,

$$C_n(x, y) = \sum_{\lambda=400}^{700} E(x, y; \lambda) r(x, y; \lambda) Q_n(\lambda) \Delta\lambda \quad (1)$$

where $Q_n(\lambda)$ is the spectral sensitivity of the n th sensor and $E(x, y; \lambda)$ is the spectral power distribution of the illumination impinging on the surface. Equation (1) means that if the number of channels n is small, with just a few broadband filters, the spectral recovery, either $r(x, y; \lambda)$ or $E(x, y; \lambda)$, represent an ill-posed problem. Many multispectral-imaging methods exploit the underlying smoothness of signal spectra, with illuminants and spectral reflectances represented by low-dimensional models based on principal component analysis (PCA) or independent component analysis (ICA) [2, 11, 12]. Thus, given a linear model, if the number of PCA (or ICA) coefficients of a particular set of spectra is the same as the number of camera responses (three in the simple trichromatic case), then the spectra can be derived by an inverse transformation of the set of camera responses, with the forward transformation being estimated from a representative (“training”) data set. If the number of coefficients is more than the number of response values, then the latter may need to be increased by imaging the scene under different illuminants or by introducing colored filters one at a time in front of the camera to modify the sensor spectra. An alternative to the

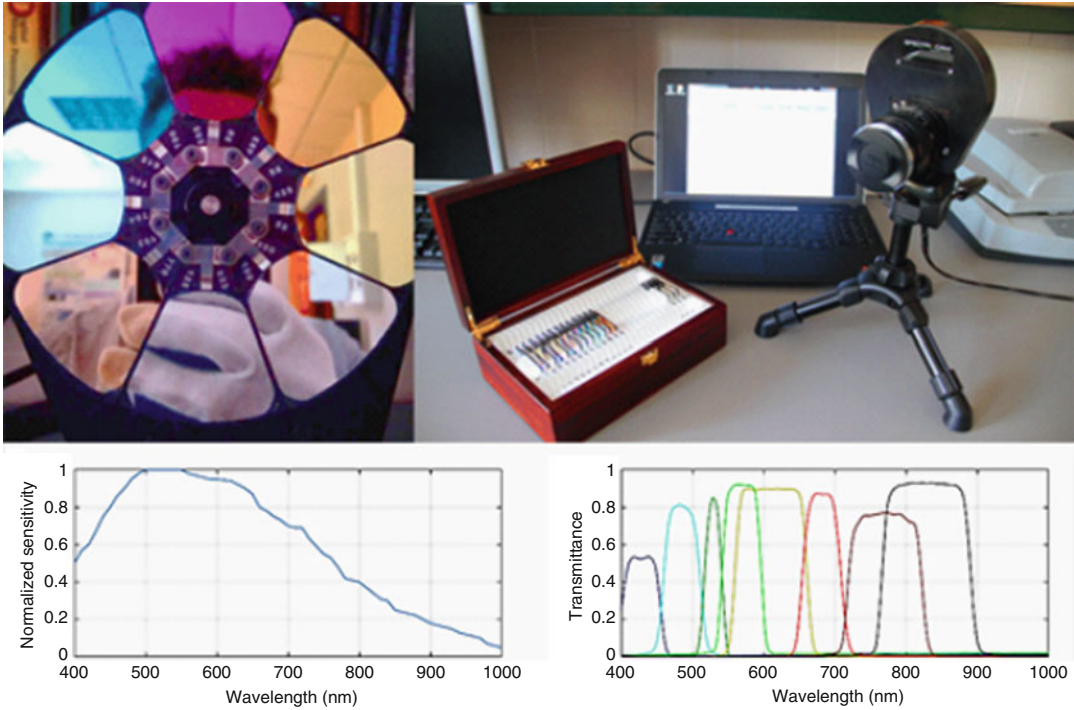
two previously mentioned configurations is the use of a monochrome camera coupled to a monochrome sensor and a set of band-pass filters (Fig. 5).

Applications

Hyperspectral imaging is used in many fields where applications demand high spatial and spectral resolution and need accurately color coding between similar colors or materials. For instance, in agriculture, hyperspectral imaging enables monitoring of plant health, pest control, and evaluation of ripeness. Classical applications can be also found in astronomy where ultraspectral imaging systems allow to inspect planets at distance or to determine a spatially resolved spectral image of stars. In food processing and research, hyperspectral imaging systems are used to identify defects, apply intelligent food classification, and characterize product quality. In remote sensing, spectral systems (when mounted in a drone) are now able to track forest health, water quality, or potential sources of fires. In general, spectral imaging is also becoming more and more popular in museums, geology, chemical imaging, skin care, cosmetics, etc.

Hyperspectral Imaging of Pictorial Artworks

Spectral imaging is a very powerful technique which, when properly applied, can retrieve useful information from art objects in a noninvasive manner. This is especially important when the objects are very old and delicate, such as paintings, ancient documents, old maps, etc. [1, 6, 13]. All these items are subject to the action of degradation agents as well as the intervention of conservation and restoration professionals. The main problem in these spectral applications is that the illumination is no longer so well controlled. Even if standard light sources are used to illuminate the painting, it will not be homogeneously illuminated, presenting highlights and shadowy areas due to the common presence of high reflectivity binders, varnishes, or materials. Moreover, different spatial locations and spectral bands might register huge variations in intensity



H

Hyperspectral Imaging, Fig. 5 Pixelteq Spectrocam VIS-NIR with the 8-slot filter wheel facility and the available set of 16 band-pass filters in the visible and near-infrared spectrum (from <https://pixelteq.com/>)

of the reflected light from the painting. This might exceed the dynamic range of the device, making it necessary to use HDR techniques to be able to recover useful information for each pixel and every spectral band [6, 13].

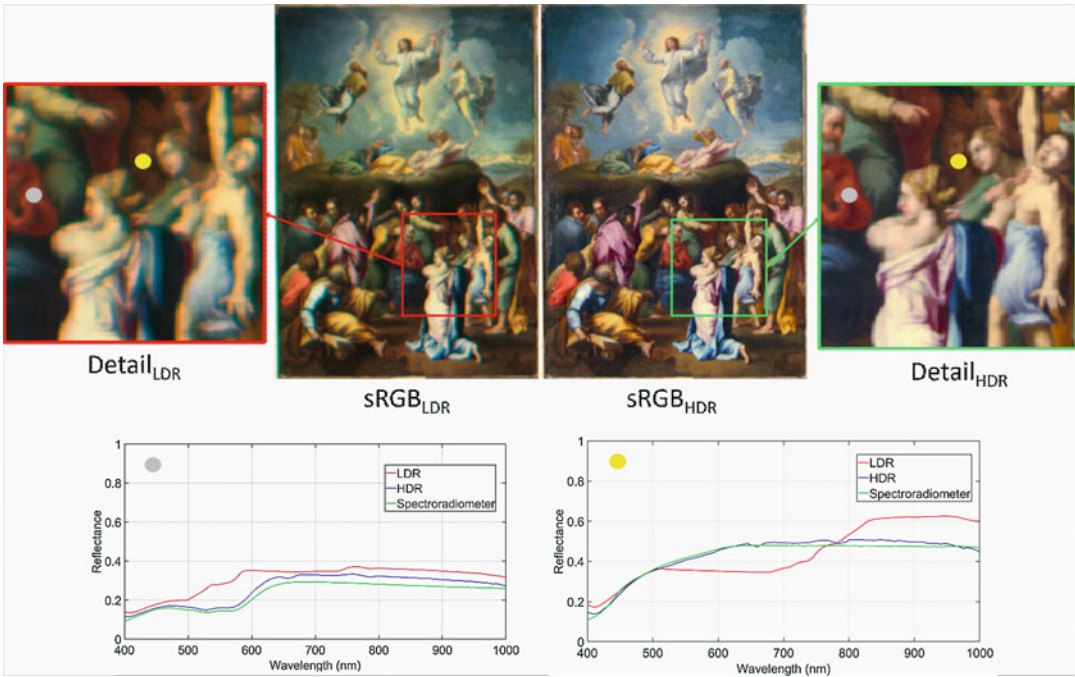
Unfortunately, hyperspectral scanners only allow one exposure time to be set for each complete capture, which means that in a single scan, all the wavelengths and areas of the scene are captured using the same single exposure time. In these uncontrolled conditions, the high dynamic range and the focusing problem due to chromatic aberrations can be overcome by using multiple captures with different focus positions and exposure times. Figure 6 shows how the use of a low dynamic range hyperspectral cube can affect the color fidelity of the rendered painting in comparison with the use of an HDR cube.

In addition, recent studies have shown how spectral imaging also allows surface geometry recovery of the whole scene. A six-band scanner can be used as a multispectral imaging system in such a way that the surface properties of a

painting, spectral reflectance of each pixel and the shape information of normal surface and height can be estimated [14]. This is of particular interest in painting analysis because experts can have access to the usually complicated shape of surfaces of paintings (i.e., irregular canvas, not uniform painter strokes on the canvas, etc.).

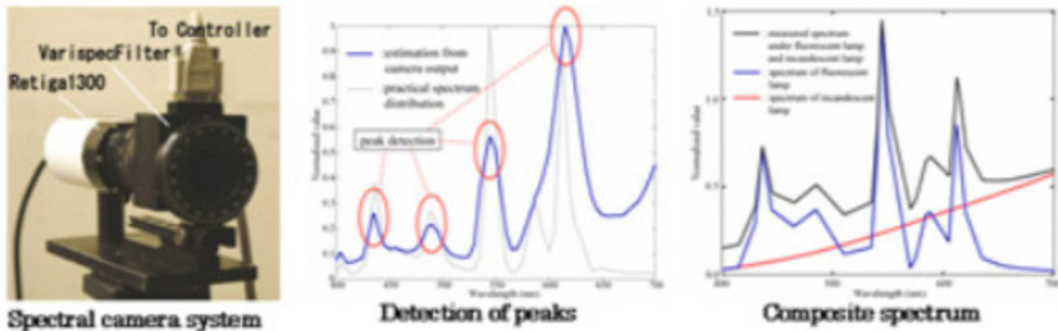
Multispectral Imaging of Natural Scenes

Acquiring outdoor spectral image data is an extra challenge because the radiometric conditions of the image scene are generally not static and the illumination does not have a smooth spectrum. Therefore, additional measures for dealing with these issues are required. Due to the great deal of time needed for hyperspectral imaging techniques, multispectral imaging based on indirect measurements can be an easy alternative. Using a conventional RGB color camera with a limited set of one to three broadband color filters (i.e., number n of bands of 4–6) in combination with a direct-mapping method [11], the spectral reflectances of natural scenes can be recovered with



Hyperspectral Imaging, Fig. 6 Comparison between a color image rendered from a hyperspectral image capture using (left) a low dynamic range LDR cube and (right) a high dynamic HDR cube. Two spectra from different areas

of the image are shown below by comparing the spectra obtained from the hyperspectral imager and conventional spectroradiometer measurements



Hyperspectral Imaging, Fig. 7 Examples of recovery of the spectral power distribution of fluorescent illuminants using a multispectral imaging system (RGB color camera with a broad-band color filter) [15]

good spectral and color quality. Given a set of training spectra $\mathbf{S}$ (i.e., spectral radiances or spectral reflectances at a pixel) and the corresponding set of camera responses $\mathbf{C}$ (as defined by Eq. 1), a recovery transformation matrix $\mathbf{D}$ can be defined by $\mathbf{D} = \mathbf{S}\mathbf{C}^+$, where $\mathbf{C}^+$ is the pseudo-inverse of $\mathbf{C}$. If $\mathbf{C}$ has full rank, then $\mathbf{C}^+ = (\mathbf{C}^t\mathbf{C})^{-1}\mathbf{C}^t$, where $\mathbf{C}^t$ is the transpose of $\mathbf{C}$. An estimate $\mathbf{S}_e$ of a set of test

spectra $\mathbf{S}_p$ may then be obtained from the corresponding set of camera responses $\mathbf{C}_p$ by applying the transformation $\mathbf{D}$, that is, $\mathbf{S}_e = \mathbf{D}\mathbf{C}_p$. Direct-mapping is an easy and a fast procedure to get spectra recovery of a pixel without consuming much time; however, the training step should be carefully taken into account to get accurate results [8, 11, 12].

Multispectral Image Capture of Natural and Artificial Illumination

Although intense research has focused on the analysis of multispectral data to recover spectral reflectance functions of objects, spectral recovery in the context of the spectral power distribution of artificial illumination has received little attention. Natural and artificial scene illuminant recovery can also be addressed by estimating the spectrum distribution of an unknown illuminant using a spectral camera system. Spectral characterization of both daylight and artificial illumination is important in fundamental research and applications, in particular for the analysis and synthesis of scenes. Nevertheless, artificial illuminants, such as fluorescent light, are characterized by complicated spectral power distributions (i.e., containing several pronounced peaks) which so far are difficult to spectrally estimate [15]. An accurate spectral estimation of fluorescent lights can be obtained by using a direct-based spectral imaging setup with a monochrome CCD camera attached to a liquid crystal tunable filter (Fig. 7).

Cross-References

- [CIE Chromaticity Coordinates \(xyY\)](#)
- [Color Appearance](#)
- [Image Quality](#)

References

1. Hardeberg, J.Y.: Acquisition and reproduction of color images: Colorimetric and multispectral approaches. Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications, pp. 121–174 (1999)
2. Imai, F.H., Berns, R.S.: Spectral estimation using trichromatic digital cameras. In: International symposium on multispectral imaging and color reproduction for digital archives, pp. 42–49. University of Chiba, Chiba (1999)
3. Hill, B.: High quality color image reproduction: the multispectral solution. In: 9th international symposium on multispectral colour science, pp. 1–7. The Society for Imaging Science and Technology, Taipei (2007)
4. Fairchild, M.D.: Color appearance models, 2nd edn. Wiley, Chichester (2013)
5. Foster, D.H., Amano, K.: Hyperspectral imaging in color vision research: tutorial. *J. Opt. Soc. Am.* **A36**(4), 606–627 (2019)
6. Hagen, N., Kudenov, M.W.: Review of snapshot spectral imaging technologies. *Opt. Eng.* **52**(9), 090901 (2013)
7. Eckhard, J., Eckhard, T., Valero, E.M., Nieves, J.L., Garrote, E.: Outdoor scene reflectance measurements using a Bragg-grating-based hyperspectral imager. *Appl. Opt.* **54**(13), D15–D24 (2015)
8. Foster, D.H., Amano, K., Nascimento, S., Foster, M.J.: Frequency of metamerism in natural scenes. *J. Opt. Soc. Am. A* **23**(10), 2359–2372 (2006)
9. López-Álvarez, M.A., Hernández-Andrés, J., Romero, J.: Developing an optimum computer-designed multispectral system comprising a monochrome CCD camera and a liquid-crystal tunable filter. *Appl. Opt.* **47**(24), 4381–4390 (2008)
10. Gao, L., Kester, R.T., Hagen, N., Tkaczyk, T.S.: Snapshot image mapping spectrometer (IMS) with high sampling density for hyperspectral microscopy. *Opt. Express* **18**(14), 14330–14344 (2010)
11. Nyström, D.: Reconstructing spectral and colorimetric data using trichromatic and multi-channel imaging. In: 9th international symposium on multispectral colour science, pp. 45–52. The Society for Imaging Science and Technology, Taipei (2007)
12. Valero, E.M., Nieves, J.L., Nascimento, S.M.C., Amano, K., Foster, D.H.: Recovering spectral data from natural scenes with an RGB digital camera and colored filters. *Color Res. Appl.* **32**(5), 352–360 (2007)
13. Martínez, M.A., Valero, E.M., Nieves, J.L., Blanc, R., Manzano, E., Vilchez, J.L.: Multifocus HDR VIS/NIR hyperspectral imaging and its application to works of art. *Opt. Express* **27**(8), 11323–11338 (2019)
14. Tominaga, S., Nakamoto, S., Hirai, K., Horiuchi, T.: Estimation of surface properties for art paintings using a six band scanner. *J. Int. Colour Assoc.* **12**, 9–21 (2014)
15. Tominaga, S., Haraguchi, H.: Estimation of fluorescent scene illuminant by a spectral camera system. In: Proceedings of SPIE, Color Imaging X: Processing, Hardcopy, and Applications, vol. 5667, pp. 128–135. San Jose (2005)

Ibn Rushd (Averroes)

Eric Kirchner

Kirchner Publications, Leiden, The Netherlands



Abu'l-Walid Muhammad bin Ahmad Ibn Rushd (Latinized name: Averroes) was born in Cordoba (Spain, 1126) and died in Marrakech (Morocco, 1198). His works range from philosophy, astronomy, and medicine to religion. He would be the most influential of Aristotle's medieval commentators, writing comments, and corrections on all the

Aristotelian works available to him [1]. Scholars in Christian Europe would refer to him as “the Commentator [of Aristotle].” The revival of Aristotelism in twelfth century Europe was mainly based on Ibn Rushd's commentaries. But in the Islamic world, Ibn Rushd's defense of rationalist philosophy would be less influential than the religious Asharite philosophy that emphasized divine manifestations in nature, as advocated by al-Ghazali.

Ibn Rushd tried to formulate a theory of light and color, as consistent with Aristotle and his followers as possible. Like Ibn Sina had done before, Ibn Rushd provided many arguments why the so-called extramission theories from Ptolemy and Euclid are absurd, and vision cannot be caused by visual rays emerging from the eye. And similar to Ibn Sina, Ibn Rushd explained vision in terms of Forms transmitted from the visible object to the eye, i.e., an intromission theory [2].

Ibn Rushd maintained that colors exist even when they are not perceived, and that light is necessary for colors to be visible [3]. Like Ibn al-Haytham had done before, Ibn Rushd considered light as playing an active role in color vision, thus breaking with the traditional Aristotelian view. But unlike Ibn al-Haytham, Ibn Rushd did agree with Aristotle that also the medium played an active role in color vision. Apparently, Ibn Rushd was not aware of the theory on vision that had been proposed by Ibn al-Haytham, successfully combining the various classical theories into one mathematical–physical intromission theory of light, color and vision.

Ibn Rushd explained the different species of color as consisting of various mixtures of bodies of much or little transparency with bodies of much or little luminosity. Since every material is composed of the four elements, with only water and air being transparent and only fire being luminous, the color of a material can be attributed to the relative amounts of the elements. This explanation was later adopted by Theodoric of Freiberg (p. 172 of Ref. [4]), who explained it in terms of four principles (much or little transparency, much or little luminosity) of colors.

Ibn Rushd did not agree with Ibn Sina's criticism of the Aristotelian ideas on color mixing. In his *Jawāmi' al-āthār al-'ulwiyya* (Short Commentary on the Meteorology: 74, 19–76, 18), Ibn Rushd defended the Aristotelian view by arguing that it referred to color mixing in a qualitative sense, but not in a quantitative sense [5, 6]. For example, Ibn Rushd argued that green is formed by mixing the yellow that exists in light red with the black that is in purple. From a modern point of view, this makes the approach philosophical rather than scientific.

Also in his description of the rainbow, Ibn Rushd closely followed Aristotle: the rainbow forms by reflection of sunlight on individual rain drops that reflect light and transmit color [4].

Ibn Rushd's ideas on color would be influential on later scholars. In his works on colors (*De coloribus*) and on the rainbow (*De iride*), Theodoric of Freiberg (d. 1318) makes clear that his theory on the formation of colors is based on Ibn Rushd's color theory (p. 47 of Ref. [7]; p. 11 of [8]). Theodoric literally quotes the Latin translation of Ibn Rushd's *Tractato de Sensu et Sensato* (p. 35n2 of Ref. [7]) [9]. His description of the colors of the rainbow, and also his four principles by which color is explained were all explicitly taken from Ibn Rushd's work.

References

1. Lindberg, D.C.: Theories of Vision from al-Kindi to Kepler. University of Chicago Press, Chicago (1976)
2. Sabra, A.I.: Optics, islamic. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, pp. 240–247. Scribner's sons, New York (1989)
3. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) Encyclopaedia of Islam. Brill, Leiden (1991)
4. Wallace, W.A.: The Scientific Methodology of Theodoric of Freiberg – A Case Study of the Relationship Between Science and Philosophy. University Press, Frimbourg (1959)
5. van Campen, M.: De regenboog bij de Arabieren. University of Utrecht, Utrecht MSc Thesis, 95 (1988)
6. Ibn Rushd: Short commentary: Kitab al-athar al-'ulwiyya. In: Osmaniya, D.M. (ed.) Rasa'il Ibn Rushd, vol. 4 Hyderabad, (1947). Another edition is: Abu Wafia, S.F., 'Abd ar-Raziq, S.A. (eds.) Kitab al-atar al-'ulwiyya. Epitome Meteorologica, Cairo (1994). (Latin translation: Aristotle, Aristotelis omnia quae extant opera, vols. V. and VI/2)
7. Krebs, E.: Meister Dietrich – Theodoricus Teutonicus de Vriberg – Sein Leben, seine Werke, seine Wissenschaft. Aschendorffsche Verlagsbuchhandlung, Münster (1906)
8. Würschmidt, J.: Dietrich von Freiberg – über den Regenbogen und die durch Strahlen erzeugten Eindrücke. Aschendorffsche Verlagsbuchhandlung, Münster (1914)
9. Cordubensis, A.: De sensu et sensato. In: Ledyard Shields, A. (ed.) Compendia librorum Aristotelis qui parva Naturalia vocantur, pp. 14–16. Mediaeval Academy of America, Cambridge, MA (1949)

Ibn Sahl, Abu Sa'd al-'Ala'

Eric Kirchner¹ and Seyed Hossein Amirshahi²

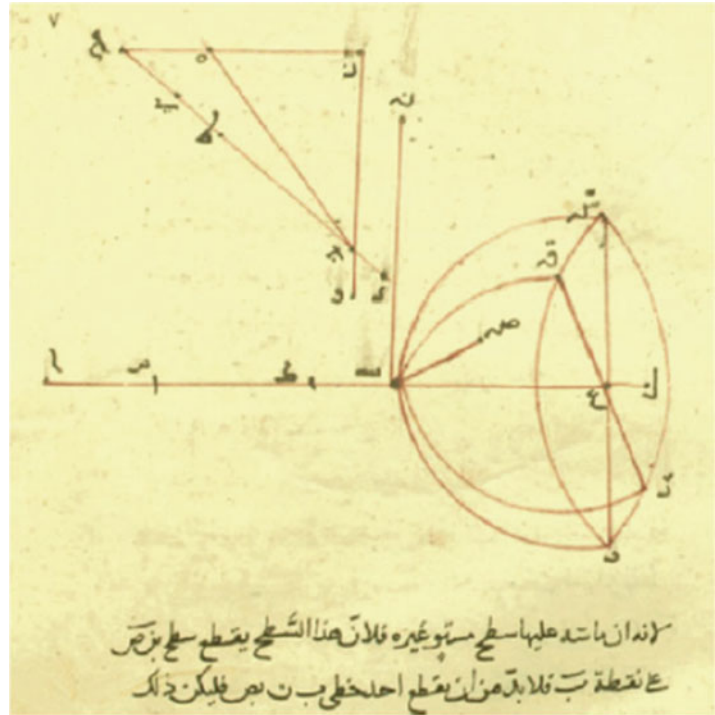
¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

Abu Sa'd al-'Ala' Ibn Sahl lived and worked as a geometer at the Abbasid (circa 940 – 1000) court in Baghdad. He wrote important works on geometric optics, mathematics, and astronomy. Nowhere in his extant works, Ibn Sahl dealt with (or even mentions) vision [1].

While investigating the transparency of the heavenly spheres that occur in Aristotelian cosmology, Ibn Sahl decided to study Ptolemy's classical work *Optics*, written in the second century AD. This made Ibn Sahl to be the first in the Arabic sources to have read and correctly understood Ptolemy's theory of refraction [1]. Ibn Sahl utilized this theory in an entirely original way for

Ibn Sahl, Abu Sa'd al-'Ala', Fig. 1 With the triangles in the top diagram, Ibn Sahl illustrated the law that is currently known as Snell's law. © Iranian National Library, Tehran. Manuscript MS 687



constructing burning instruments such as lenses and glass spheres by means of refraction.

In the year 984, Ibn Sahl wrote the treatise "On Burning Mirrors and Lenses." In this work, he investigated the optimum (nonspherical) shape of lenses and mirrors to focus light at a given distance. Ibn Sahl "appears to be the first in history to engage in research on burning lenses" [2]. He subsequently treated the parabolic mirror, the ellipsoidal mirror, the plano-convex lens, and the biconvex lens. His calculations on geometric aberration predate similar calculations done by Descartes in the 1620s. Ibn Sahl carried out these calculations both for light sources at an (almost) infinite distance such as the sun and for light sources at finite distances [3].

In these calculations, Ibn Sahl needed a law of refraction. Intriguingly, for this he used a law that is geometrically equivalent to Snell's law that would be rediscovered in 1621 by Willebrord Snellius and in 1602 by Thomas Harriot. The sine law of refraction was thus discovered by Ibn Sahl [4, 5].

The illustration shows a page from Ibn Sahl's manuscript. In the top part of the figure, two

overlapping triangles are shown. The hypotenuse of the external triangle represents the direction of incident light, while the hypotenuse of the internal triangle represents the direction of refracted light inside a transparent medium. By demanding that these two direction vectors intersect in one point, constructing the direction of refracted light with this figure is geometrically equivalent to keeping the ratio of sines of incident and refracted angles constant. This is Snell's law of refraction.

Ibn Sahl used this law of refraction several times in the treatise, but without explicitly stating it as a law [4, 5]. Indeed, the concept of natural law did not exist at the time. Ibn Sahl used his law as if it was a mathematical relation that was well known. He repeatedly applied this relation, utilizing the fact that the ratio between the sines of incoming and refracted angles is constant. He made no reference to the fact that this constant depends on material-dependent properties, i.e., what is now known as their refractive index.

Ibn Sahl's treatise was later used by Ibn al-Haytham in his investigations of refraction. Interestingly, Ibn al-Haytham apparently did not

recognize the law of refraction as used by Ibn Sahl. Instead, Ibn al-Haytham started his own experimental investigation into finding a law of refraction.

References

1. Sabra, A.I.: Ibn al-Haytham's revolutionary project in optics: the achievement and the obstacle. In: Hogendijk, J.P., Sabra, A.I. (eds.) *The Enterprise of Science in Islam*, pp. 85–118. MIT Press, Cambridge, MA (2003)
2. Rashed, R.: Geometrical optics. In: Rashed, R. (ed.) *Encyclopedia of the History of Arabic Science*, vol. 2, pp. 643–671. Routledge, London (1996)
3. Berggren, L.: Ibn Sahl: Abu Sa'd al-'Ala'ibn Sahl. In: Hockey, T. (ed.) *The Biographical Encyclopedia of Astronomers*, p. 567. Springer, New York (2007)
4. Rashed, R.: *Géométrie et dioptrique au Xe siècle: Ibn Sahl, al-Quhi et Ibn al-Haytham*. Les Belles Lettres, Paris (1993)
5. Rashed, R.: A pioneer in anacastics: Ibn Sahl on burning mirrors and lenses. *Isis*, **81**, 464–491 (1990)

ICC L*a*b*

- [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

ICCLAB

- [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

Ideal Black Surface

- [Blackbody and Blackbody Radiation](#)

Ideal Emitter

- [Blackbody and Blackbody Radiation](#)

Ideal Observer Models of Color Category Learning

- [Bayesian Approaches to Color Category Learning](#)

Ideal Thermal Radiator

- [Blackbody and Blackbody Radiation](#)

Illuminance Meter

Bor-Jiunn Wen¹ and Chao-Hua Wen²

¹Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Luxmeter](#)

Definition

In photometry, Illuminance Meter is a device that measures the total luminous flux incident on a surface, per unit area. Illuminance is a measure of how much the incident light is spread over a given surface, wavelength-weighted by the luminosity function to correlate with the brightness perception of human eyes.

Introduction

There are many units for defining illuminance. The International System of Units defines using the unit of lux (symbol: lx) or lumens per square meter (cd sr m^{-2}), which is most frequently used. In the centimeter-gram-second system, the unit of

illuminance is the phot, which is equal to 10,000 lx. The foot-candle is a nonmetric unit of illuminance that is also used in photography. Illuminance is sometimes called brightness, but this leads to confusion with other uses of the word, such as brightness perception. Hence, “brightness” should never be used for quantitative description, but rather for describing the visual perceptions of light or color.

The human eye is capable of seeing within a range of more than two trillion-fold. The presence of white objects is somewhat discernible under starlight, at 5×10^{-5} lx, while at the bright end, we can easily read large text at 10^8 lx, or at about 1000 times that of direct sunlight, although this can be very uncomfortable and cause long-lasting after-images.

Application 1

In addition to an illuminance meter, illuminance color spectral meters including spectrometer, spectrophotometer, spectrograph, or spectroscope are also important for measuring properties of light over a specific portion of the electromagnetic spectrum, typically used in spectroscopic analysis to identify materials [1]. The parameter measured is most often the light’s intensity, but could also be, for instance, the polarization state. The independent variable is usually the wavelength of the light or a unit directly proportional to the photon energy, such as wave number or electron volts, which has a reciprocal relationship with wavelength. A spectrometer is used in spectroscopy for producing spectral lines and measuring their wavelengths and intensities. “Spectrometer” is a term applied to instruments that operate over a very wide range of wavelengths, from gamma rays and X-rays into the far infrared. If the instrument is designed to measure the spectrum in absolute units rather than relative units, then it is typically called a spectrophotometer. The majority of spectrophotometers are used in spectral regions near the visible spectrum. A spectrograph is an instrument that separates an incoming wave into a

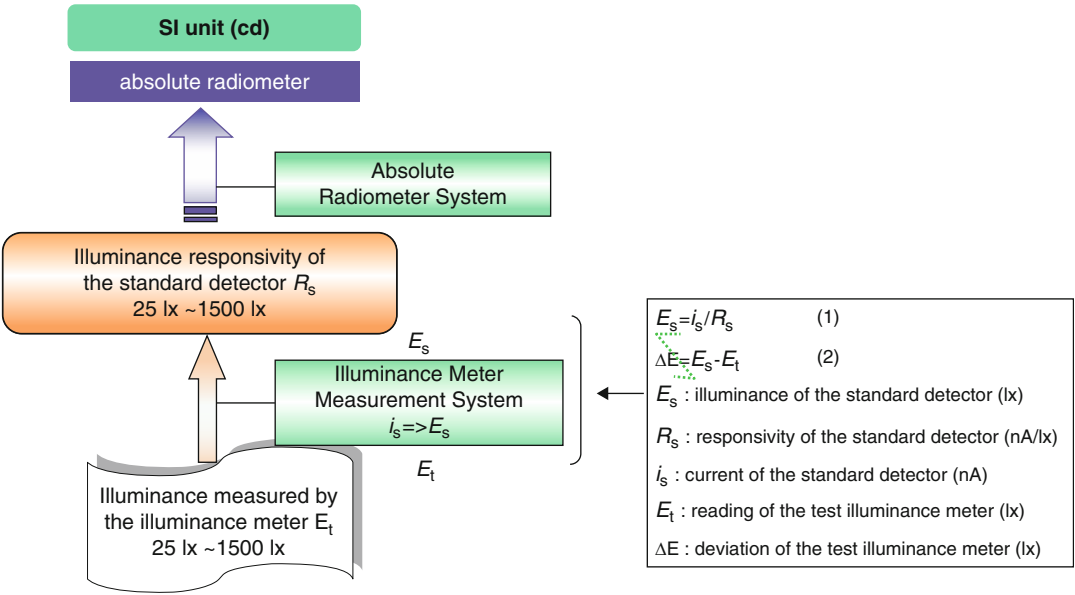
frequency spectrum. There are several kinds of machines, referred to as spectrographs, depending on the precise nature of the waves. Spectroscopes are often used in astronomy and some branches of chemistry. Early spectroscopes were simply prisms with graduations marking wavelengths of light. Modern spectroscopes generally include a diffraction grating, a movable slit, and some kind of photodetector. It is all automated and controlled by a computer.

Application 2

In general, much of computer vision, image processing, and imaging is predicated on the assumption that there is a single prevailing illuminant lighting a scene. However, often multiple lights are present. Common examples include outdoor scenes with cast and attached shadows, indoor office environments that are typically lit by skylight and artificial illumination, and the spotlighting used in commercial premises and galleries. Relative to these mixed lighting conditions, many imaging algorithms based on the single light assumption can fail. Therefore, Finlayson et al. [2] proposed a method of detecting multiple illuminants in images based on the chromogenic theory. By forcing the lights to be examined pair-wise without insisting on the accuracy of the illuminant estimation, and by processing the results on the basis of region rather than pixels, the research obtained very accurate results over a variety of illuminations. Accordingly, the results are remarkably good, especially considering the simplicity and speed of the method.

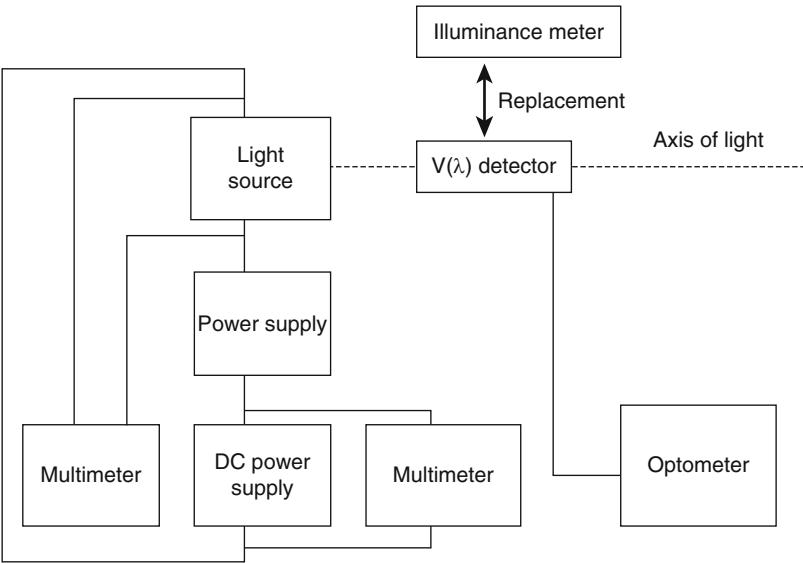
Calibration

In the color measurement field, calibration is desired to ensure that the system is closely monitored in well working conditions. The traceability chain of illuminance is presented in Fig. 1 [3]. An illuminance meter calibration set-up is shown in Fig. 2 [3]. First, warm up the multimeters. Check



Illuminance Meter, Fig. 1 Traceability chain of illuminance

Illuminance Meter, Fig. 2 Schematic diagram of illuminance meter calibration



that the power of the device under test (DUT) is sufficient. Check the connections of all the instruments as shown in Fig. 1. They should be well connected, including the polarity of the lamp. Second, align the lamp filament to the axis of the light source and fix the position of the DUT.

Third, turn on the power supply after the multi-meters have already warmed up. Ramp up the lamp current to operation value. Execute calibration after the standard lamp has stabilized. Fourth, read the output current of the $V(\lambda)$ [4] (the spectral luminous efficiency function for photopic vision)

detector, and calculate the illuminance of the position of the $V(\lambda)$ detector. Finally, replace the $V(\lambda)$ detector with the test sample at the same position, and record the test reading on the recording card.

Cross-References

- [Instrument: Spectrophotometer](#)
- [Melanopsin Contributions to Human Brightness Perception](#)
- [Spectrometer](#)

References

1. Butler, L.R.P., Laqua, K.: Nomenclature, symbols, units and their usage in spectrochemical analysis – IX. Instrumentation for the spectral dispersion and isolation of optical radiation. *Pure Appl. Chem.* **67**(10), 1725–1744 (1995)
2. Finlayson, G., Fredembach, C., Drew, M.S.: Detecting illumination in images. In: *Computer Vision, 2007. IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–8, 14–21 Oct 2007
3. Chung, T.Y.: *Instrument Calibration Technique for Illuminance Meter of Absolute Radiometer System*, 07-3-80-0086, 10th edn. Center for Measurement Standards/Industrial Technology Research Institute (2019)
4. Ohta, N., Robertson, A.R.: *Colorimetry Fundamentals and Applications*. Wiley, Chichester (2005)

Illusory Colors

- [Fechner's Colors and Benham's Top](#)

Image Burn-in

- [Afterimage](#)

Image Fidelity

- [Image Quality](#)

Image Quality

Herzog Robert
Max-Planck Institute for Informatics,
Saarbrücken, Germany

Synonyms

[Image fidelity](#); [Image similarity](#)

Definition

Image quality is commonly characterized as the perceived image degradation with respect to an ideal undistorted image.

Overview

For many applications in research and industry, there is a constant need for quality assessment of images (e.g., computer graphics, image compression, camera manufactures, medical imaging).

Image quality cannot be formalized in general since it plays a different role depending on the application. For example, in *lossy compression* and *streaming*, the perceived quality-to-bit rate ratio with respect to a reference image needs to be maximized, whereas in computer vision, forensic, and medicine, the image quality is driven by *task performance* (i.e., how much semantic information is conveyed in the image). In photo-realistic image synthesis (e.g., 3D computer games, movies), image quality can be regarded as a measure of realism (photography versus rendered image). Moreover, in art and in particular in photography, image quality is strongly influenced by content and salience (see ► [“Image Fidelity”](#)).

Apart from photographic lens distortions, image quality is largely affected by common factors such as noise, blurring, ringing (sharpening), aliasing, dynamic range compression, color quantization, image resolution, and application-specific artifacts (e.g., JPEG compression, rendering artifacts).

Image Quality Assessment

Image quality assessment can be divided into *subjective quality assessment* and *objective quality assessment*.

In subjective quality assessment, human observers independently and subjectively quantify the quality of an image by assigning it a single number that represents the quality score. The average is referred to as the *mean opinion score* (MOS). Objective quality assessment refers to the task of automatically evaluating the quality of images in a way human observers would do.

With today's new technologies, the spectrum of image quality degradations has widened including distortions coming from 3D stereoscopy, ► [high-dynamic-range \(HDR\) imaging](#), 3D image synthesis, and light-field imaging.

Image Quality Metrics

Automatic image quality metrics (IQM) have been developed for the task of objective image quality assessment. Although primarily designed for analyzing still image quality, IQMs have also been exploited for the task of video quality control. However, the human perception of video quality requires proper spatiotemporal modeling of the human visual system (HVS) including also the perception of temporal artifacts like flickering, which is covered by special *video quality metrics*. IQMs can be classified as *full-reference*, *reduced-reference*, and *no-reference* image quality metrics (the latter often referred to as *blind IQM*).

Full-Reference Quality Metrics

Full-reference image quality metrics aim at measuring the strength of general image distortions given two corresponding images, a distorted and a reference (undistorted) one. The given reference image simplifies the detection of universal distortions and their spatial localization in the image but, on the other hand, requires a relatively precise alignment of the two images.

When predicting a single image quality score similar to the subjective MOS, a two-stage approach is pursued. First, a local distortions measure is computed to obtain a quality map. Second,

a spatial pooling algorithm is employed to integrate the local distortions into a single number. Most IQM ignore the *chroma* component and focus only on *luminance* for which the human visual system is most sensitive to.

Full-reference quality metrics can be further divided into *physical IQMs*, *perceptual IQMs*, and *structural and information-theoretic IQMs*.

Physical IQMs have a clear physical meaning, are fast to compute, and are easy to differentiate making them preferred candidates for mathematical optimization techniques (energy minimization). Common physical IQMs are *mean-square error* (MSE), *peak signal-to-noise ratio* (PSNR), *mean absolute error*, and its generalization the *Minkowski error metric* using the l_p norm:

$$E_M = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{1/p}$$

where x_i, y_i are the values of the i th pixel in the distorted and reference image respectively and p is a positive real number.

Physical IQMs consider an image as a collection of N independent pixels x_i thereby ignoring any correlation among them. As shown in Fig. 1, the human-perceived error with respect to a reference image varies locally depending on the brightness, frequency, and correlation of the reference image content and the error signal.

Therefore, *perceptual image quality metrics* have been developed in order to better predict the perceived image quality (image similarity), which they achieve by simulating the early pathways of the human visual system (HVS) modeling effects such as *light adaptation*, *luminance masking* (Weber's law), *contrast sensitivity function* (CSF), and *contrast masking* (also known as *visual masking*). For this reason they are also referred to as *bottom-up* quality measures. Various perceptual models for image quality were proposed: the visual difference predictor (VDP) [5], the Lubin model [3], HDR-VDP [2], Watson's DCT model [4], JPEG 2000 [6], etc. Many of these models are relatively complex and computationally intensive but predict the human-perceived distortions reliably near the visibility

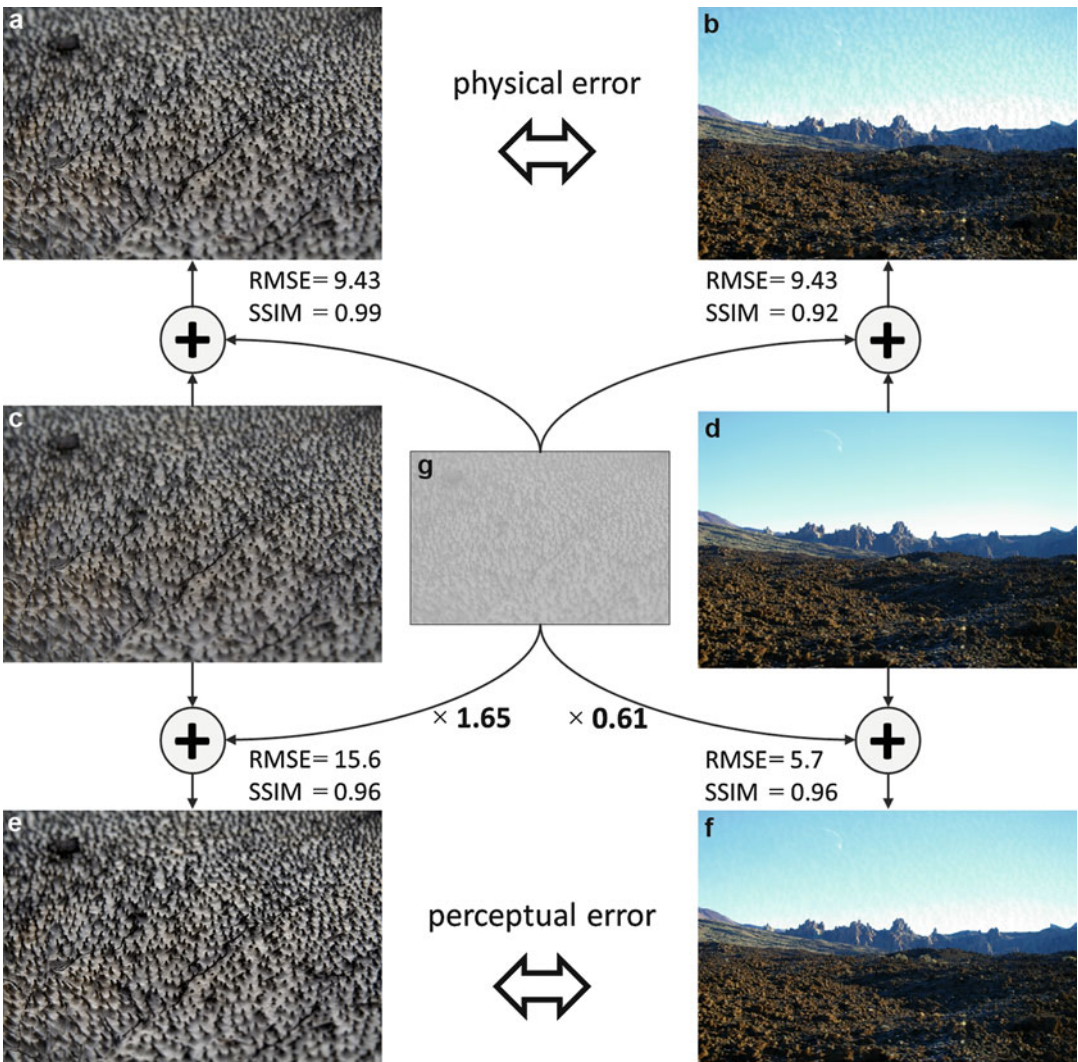


Image Quality, Fig. 1 Nonlinearity of the perception of image quality: images (a) and (b) have exactly the same root-mean-square error (RMSE) with respect to the original images (c) and (d), respectively. However, the distortions in image (b) are visually disturbing while they are almost invisible in image (a). This perceptual effect is referred to as *masking*. Note that the distortions in image (b) are more

salient in dark regions and regions with frequency content (e.g., sky) different from the error signal (g). Images (e) and (d) exhibit approximately the same perceived image quality as predicted by the structural similarity (SSIM) index metric [7]. Please note that the distortions perceived in the images depend on the display size of the images

threshold (super-threshold). However, because they are mostly tuned to experimentally acquired visibility thresholds measured on simple stimuli, the prediction of strongly visible distortions (supra-threshold) is questionable and remains still an open research field.

To overcome the complexity of HVS-based metrics, structure-based metrics (e.g., SSIM [7])

have been derived that do not explicitly model the HVS but merely compute a relatively simple statistical measure of image similarity based on local pixel-value correlation and contrast.

Information-theoretic image quality metrics predict the perceived image quality using a statistical approach. They estimate how much information is shared between the reference and distorted

image assuming the distorted image is transmitted over a noisy communication channel. Exploiting entropy-based methods known from signal coding and compression, the perceived quality of an image is computed from the *mutual information* of the reference and distorted image. A known information-theoretic quality metric is, for example, the *visual information fidelity* (VIF) index [8].

No-Reference Image Quality Metrics

No-reference (blind) image quality metrics pursue a top-down approach for predicting the quality of a single image. Thereby, the HVS is treated as a black box, and the quality of an image is predicted utilizing classification or regression techniques that operate on distortion-specific image features (e.g., gradient histograms, wavelet coefficients). No-reference IQMs require careful modeling of the image features and need many image examples for training a supervised learning algorithm (e.g., a support vector machine) [9]. Therefore, no-reference IQMs do not generalize well and are typically tuned to detect only one or a few specific image distortions (e.g., JPEG blockiness, noise, blur) where the success of the prediction primarily depends on the complexity of the type of distortion. Typical no-reference IQMs are based on features computed from Natural Image Statistics.

Reduced-Reference Image Quality Metrics

In a streaming scenario, full-reference IQMs cannot directly be applied for predicting the image quality on the client side since only the server knows the reference and the distorted image. On the other hand, no-reference IQMs are less reliable and in general also more expensive to compute than full-reference IQM. Therefore, a new paradigm for quality control has emerged, which is called reduced-reference IQM. Reduced-reference IQMs sparsely encode auxiliary information about the reference image that is streamed together with the lossily encoded image and enables the client to faithfully compute a quality score for the decoded image. The more data about the reference image is streamed, the better the

client can predict the image quality degradation. Thus, reduced-reference IQMs trade quality information for bit rate and can be regarded as a compromise between full-reference IQM and no-reference IQM.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Image Fidelity](#)
- [Light Distribution](#)
- [Luminance Meter](#)

References

1. Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. Morgan & Claypool (2006)
2. Mantiuk, R., Kim K.J., Rempel A.G., Heidrich W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* **30**, 40:1–40:14 (2011)
3. Lubin, J.: A visual discrimination model for imaging system design and evaluation. In: *Vision Models for Target Detection and Recognition*, pp. 245–283. World Scientific (1995)
4. Watson, A.: DCT quantization matrices visually optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV*, **1913**, 202–216 (1993)
5. Daly, S.: The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In: *Digital Image and Human Vision*, pp. 179–206. MIT Press, Cambridge, MA (1993)
6. Zeng W., Daly, S., Lei, S.: Visual optimization tools in JPEG 2000. In *IEEE International Conference on Image Processing*, 37–40 (2000)
7. Wang, Z., Bovik, A.C., Sheik, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. **13**, 600–612 (2004)
8. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. In *IEEE Transactions on Image Processing*. **15**, 430–444 (2006)
9. Cortes, C., Vapnik V.: Support-vector network. In *Machine Learning* **20**, 273–297 (1995)

Image Similarity

- [Image Quality](#)

Impressionism

Georges Roque
Centre National de la Recherche Scientifique
(CNRS), Paris, France

Definition

Impressionism is considered as the first vanguard movement of Occidental art history; it opens modernity in art. Its starting point is a group of artists who had in common a rejection of academic art. They organized eight collective exhibitions between 1874 and 1886. Most of the artists only took part in some of these exhibitions, like Cézanne. The painters considered as more “typically” impressionist are above all Claude Monet, Pierre-Auguste Renoir, and Camille Pissarro.

Overview

The term “impressionist” was coined by a critic as a nickname in order to criticize the movement, as he was struck by a painting presented by Monet in the first 1874 collective exhibition and entitled *Impression, soleil levant* (*Impression, Sunrise*). The novelty of Impressionism is twofold. First, it has to do with the subject matter, often taken from everyday life (train stations, boulevards, scenes of leisure, etc.), corresponding so to Baudelaire’s conception of modernity. And second, it is technical: the way of rendering light, color, atmosphere, the nature of the brushstrokes, the innovative composition, and the impression of unfinished and incompleteness are among the main features of Impressionism.

Theory and Practice

According to the standard account of Impressionism (largely widespread by one of his best

specialists, John Rewald), the Impressionists did not need any theory, since they trusted their eye. For sure, some of the statements made by Monet and Renoir seem to reinforce such a conception. However, the issue of the relationship between Impressionism and color science is more complex. The first part of this paper is devoted to general features of Impressionism regarding color and the second one to the conception of color that can be found in Monet, Renoir, and Pissarro.

First of all, it must be noted that, unlike Neo-impressionist painters, the Impressionists did not write “theoretical” essays, so that they have not conveyed a general idea of their principles. Furthermore, what they had in common was a rejection of academic art more than a clear doctrine. For this reason, the practice of the group members differs a lot from each other and depends more on personal skills than on shared beliefs, even though they had in common some general aims, like the ideal of modernity or the fact of preferring the rendering of ever-changing moments and atmospheres instead of permanent features. Yet, this does not mean that they had an “empirical” practice. Even though Monet liked to say, “I have already had a horror of theories” [1, p. 586], it should be recalled that such a statement does not mean that Monet rejected all theories but only dogmatic submission of practice to a theory prior to it, as did Bonnard and Matisse, too.

An often-quoted passage is usually given as an example of Monet’s “empirical” practice: “When you go out to paint, try to forget what objects you have before you, a tree, a house, a field or whatever. Merely think here is a little square of blue, here a oblong of pink, here a streak of yellow, and paint it just as it looks to you, the exact color and shape, until it gives your own naïve impression of the scene before you” [2, p. 35]. This passage recalls Ruskin’s famous conception of an “innocent eye”: “The whole technical power of painting depends on our recovery of what may be called the *innocence of the eye*: that is to say, of a sort of childish perception of these flat stains of colour, merely as such, without consciousness of what they signify” [3, p. 27].

Georges Roque has retired.

Monet – who praised a lot Ruskin's *Elements of Drawing* – coincided with Ruskin in the attempt to retrieving the “innocent” eye, that is to say, trying to paint things not as they are known but as they are seen. The problem, however, is that there is no “purely” visual perception: the so-called innocent eye also depends on an implicit or explicit knowledge about the perceptual ► **appearance** of what is seen, knowledge that, in turn, depends on cultural background. Not surprisingly, the fact of seeing out in the world not “objects” but stains of colors corresponds to what scientists said at the time: perception of objects is the result of a psychophysiological process in the brain [4, p. 427]. Similar ideas were already present in the field of literature and art criticism. In his obituary of Delacroix, Baudelaire wrote that there are neither lines nor colors in nature, for they are created by man [5, p. 752]. It must be added that the vocabulary of artists is also borrowed from science. From this point of view, “impression” corresponds to the empiricist view according to which in visual perception it is the first impression not yet reworked by thought. The new meaning given by Helmholtz to the concepts of “impression” and “sensation” had been popularized in France through Taine's *De l'Intelligence* (1870); artists, however, often used them as if they were equivalent.

Optical Mixture

It is generally assumed that the Impressionists achieved the particular and luminous effect of their paintings through a particular technique, known as “division of the tone” and consisting in breaking down colors into their constituents. Duranty, one of the first critics of the movement, explained indeed, “Proceeding from intuition to intuition, they [the Impressionists] have little by little succeeded in breaking down sunlight into its rays, its elements and to reconstitute its unity by means of the general harmony of the iridescent color which they spread on their canvases [...] The most learned physicist could find nothing to criticize in their analyses of light...” [2, p. 4].

It could be thought that critics were just trying to find a way to give a scientific legitimacy to a young movement highly criticized, arguing that their way of painting was “scientific,” which in fact is not true [6]. However, besides similar comments made by other critics, there is a more precise text based on conversations with Pissarro and that the painter never disowned:

About that time [1870–1871], M. Pissarro, after a holiday at Piette in Britain, went to London with M. Claude Monet. During their 10 months stay in this city, they sent canvases to the National Academy, which then condemned the painters' audacity. Nevertheless, on British soil, their eyes had been educated by Turner, who had for a long time restricted himself solely to the colors of the prism. In their study of his work, they found confirmation of theories already discussed in private and realized in individual essays that had not yet been publicly shown: *the law of complementary colors and its natural end, the division of the tone*.

This stay in England hastened the Impressionist's evolution. Back in Paris, Pissarro and Monet made themselves the exegetes of the new technique. Quickly, their friends, prepared by previous attempts, recognized the superiority of the retinal color mixture over the obviously darker mixture that occurs on the palette. The optical reconstitution of the complementary colors divided on the canvas finally gave them those blond lights so patiently sought.

Impressionism, stemming from precise theories, soon emerged with the brilliance of its luminous and vibrating harmony [7]

This text shows at least that Pissarro and Monet attempted something considerably beyond “empirical” practice of painting: they needed theoretical explanations grounded on the scientific knowledge available in their time in order to confirm their practice. Indeed, the “law of complementary colors” refers to the theory of the French chemist Michel-Eugène Chevreul, who was very influential at the time and is directly or indirectly the main source of the Impressionist color theory [8]. According to Charles Blanc, a very important art historian (he was also director of the Paris École des Beaux-Arts), whose book *Grammar of*

Painting and Engraving was avidly read by the artists at the time, “it is the mutual exaltation of juxtaposed complementary colors that Mr. Chevreul called ‘the law of simultaneous contrast of colors’” [9, p. 562]. If two ► **complementary colors** are juxtaposed, indeed, they enhance each other, and Impressionist painters were eager to use bright colors. For this reason, they took seriously the harmony of complementary colors. They also thought that they would obtain a more luminous result when juxtaposing dots of colors (and particularly dots of complementary colors) that were supposed to fuse in the eye; for this reason, the “division” of the tone is directly related to the law of complementary contrast. What has been called “indigomania,” due to a pervasive use of violet and blue in Impressionist canvases, has a lot to do with this issue [8, pp. 34–36]. Georges Guérout (one of the translators of Helmholtz into French) gave indeed an explanation for the use of violet and blue, in particular for shadows: “The *plein air* school is right to put blue or violet in the shadows. For yellow, whose complement is violet, dominates much in the sun, and, by contrast, violet dominates in the shadow” [10, p. 176].

Artificial Pigments

Besides harmony of ► **complementary colors**, another general feature of Impressionism is the use of artificial pigments that had been recently introduced in the color market [11, 12]. Furthermore, they liked to use them pure, as they come from the tube, that is, without mixing them. Such a use has been deeply criticized at the time, as pure hues were not considered as “true” artistic colors. According to this dated conception, colors, in order to be accepted as artistic, have to be prepared and mixed. The reason for the Impressionist practice regarding pigments is related to their modernism: painters were proud to use newly created synthetic pigments instead of the old ones; furthermore, using them pure was also another way of being modern, as it entailed a strong critique of the traditional view according to which painters had to prepare their own colors.

Renoir, Monet, and Pissarro

After this general overview, a look at some painters in particular could confirm, despite what is generally assumed, that they were interested in color theory and applied it to their works. From this point of view, artists’ statements do not have to be taken literally. Renoir once explained that he painted “like a child,” and he added, “I don’t have neither rules nor methods” [13, p. 13]. However, a closer look at his writing shows that he was aware of the fact that the brightness of a hue depends on the other colors contiguous to it. When Julie Manet asked him advices as how to improve a still life on a white tablecloth, Renoir explained to her that “one has to give white its intensity thanks to the value of what surrounds it, and not by adding more white to it” [13, p. 223]. Now, this is exactly one of the six advices Chevreul gave painters: white placed beside a color heightens its tone [14, § 336].

Exactly the same warning can be made about Monet’s statements. And in his case, too, relevant information confirms his interest for color theory. In an interview given in 1888 to an English newspaper, he explained, “Color owes its brightness to force of contrast rather than to its inherent qualities [...] ► **primary colors** look brightest when they are brought into contrast with their complementaries” [11, p. 88]. This statement explains his well-known taste for the field of poppies he painted very often, as they gave him an excellent opportunity to juxtapose green surfaces with patches of red, enhancing so the two ► **complementary colors**. Another source of information is the visit the painter Louis Anquetin made to Monet in order to inquire into his conception of color. According to Émile Bernard’s account of this visit, Anquetin was disappointed, realizing that Monet “only knew a very little part of what he was coming to ask him, contenting himself with the theory of the complementary colors, applied with the palette of the seven prismatic colors” [15, p. 112]. This information confirms that Monet used the harmony of complementary colors. It should also be noted that Monet’s well-known interest for light reflections in water has also to do with scientific theories [16].

Now, what about Pissarro? According to Geffroy, he “was always passionately and learnedly interested in the theories of lights and colors, fervent in the search for processes, always ready for research and realization” [17, p. 263]. Not surprisingly, his contribution to color is also remarkable, in particular concerning frames. As a critic recalled, “At the 1877 Exhibition, M. Pissarro, applying in its rigorous logic the law of the complementary colors, set his canvases into white frames that, without influencing the colors, left the tones with their exact values” [7]. Indeed, Chevreul had actually drawn attention to the influence of the colors of the frame on the colors of the framed work and had argued against the use of gilded frames [14, §565]. Since Chevreul was probably the only scientist to have paid attention to this problem at the time, Pissarro could not have learned it from any other source. Pissarro liked the idea and used white frames for his paintings, after convincing his dealer to do so. In 1881, for the Sixth Impressionist Exhibition, Pissarro innovated again: he tinted the stretchers with the complementaries of the color dominating in the painting, which was another original way of enhancing the colors of his paintings.

To conclude, even though in some statements Impressionist painters said they had a purely empirical conception of color and were not interested in theory, there is enough evidence showing that they were using recipes allowing them to enhance their colors. For this aim, Chevreul’s law of simultaneous contrast, as understood by Blanc and others, provided excellent tools and advices. Closer technical analysis of Impressionist paintings confirms their frequent use of juxtaposed ► [complementary colors](#) [11, 12].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Appearance](#)
- [Chevreul, Michel-Eugène](#)
- [Color Contrast](#)
- [Color Harmony](#)

- [Complementary Colors](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Neo-impressionism](#)
- [Primary Colors](#)

References

1. Monet, C.: Letter to Charteris (June 21, 1926), quoted by Rewald, J.: *The History of Impressionism*, 4th edn. Museum of Modern Art, New York (1973)
2. Nochlin, L.: *Impressionism and Post-Impressionism 1874–1904. Sources and Documents*. Prentice Hall, Englewood Cliffs (1966)
3. Ruskin, J.: *The Elements of Drawing*. Dover, New York (1971)
4. von Helmholtz, H.: *Handbuch der physiologischen Optik*. Leopold Voss, Leipzig (1867)
5. Baudelaire, C.: *Oeuvres complètes II*. Gallimard, Paris (1976)
6. Carson Webster, J.: The technique of impressionism: a reappraisal. *College Art J.* **IV**(1), 3–22 (1944)
7. Lecomte, G.: Camille Pissarro. *Les Hommes d’aujourd’hui* **VIII**(366) (1890)
8. Roque, G.: Chevreul and impressionism: a reappraisal. *Art Bull.* **LXXVIII**(1), 25–39 (1996)
9. Blanc, C.: *Grammaire des arts du dessin*. Renouard, Paris (1880). The chapter on color has been translated into English in Taylor, J.C. (ed.): *Nineteenth-Century Theories of Art*. University of California Press, Berkeley (1989)
10. Guérault, G.: Formes, couleurs et mouvements. *Gazette des Beaux-Arts.* **XXV**, 165–179 (1882)
11. Bomford, D., et al.: *Art in the Making. Impressionism*. The National Gallery, London (1990)
12. Callen, A.: *The Art of Impressionism. Painting Technique & the Making of Modernity*. Yale University Press, New Haven/London (2000)
13. Renoir, P.-A.: *Écrits, entretiens et lettres sur l’art*. Éditions de l’Amateur, Paris (2002)
14. Chevreul, M.-E.: *De la loi du contraste simultané des couleurs...* Pitois-Levrault, Paris (1839). The latest English translation is: *The Principles of Harmony and Contrast of Colors and their Applications to the Arts* (1855). Kessinger Publishing LLC, Whitefish (2009)
15. Bernard, E.: Louis Anquetin. In: *Gazette des Beaux-Arts* 76th Year, vol. XI(Part 2), pp. 108–121 (1934)
16. Roque, G.: Au fil de l’eau: scintillements, papillotement, miroitement, lustre, vibrations. In: *Éblouissants reflets. Cent chefs-d’œuvre impressionnistes*, exhib. cat, pp. 70–79. Musée des Beaux-Arts, Éditions de la Réunion des musées nationaux, Rouen (2013)
17. Geffroy, G.: *Monet, sa vie, son oeuvre*. Macula, Paris (1980). (1st ed. 1924)

Inappropriate Use of Color

► Color Pollution

Incandescence

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Definition

Emission of optical radiation by the process of thermal radiation.

Thermal radiation is electromagnetic radiation originating in the thermal motions of particles of matter (atoms, molecules, ions, etc.). All matter above the temperature of absolute zero emits thermal radiation. Objects around room temperature emit mainly infrared radiation. Objects at temperatures higher than some 800 K also emit electromagnetic radiation in the visible area.

Thermal Light Radiators

- The Sun that generates light by its hot plasma with a temperature of around 6,000 K
- Glowing coal and red hot horseshoe with a temperature of around 1,000
- Torches, oil lamps, candles, and gas lamps with temperatures of around 2,000
- Carbon-arc lamps with temperatures of up to 5,000 K
- Old-fashioned chemical flash bulbs with temperatures of around 4,000 K
- Incandescent lamps with temperatures of 2,700 K
- Halogen incandescent lamps with temperatures of 2,900–3,200 K
- Black body radiator

Black Body Radiator

A black body radiator is an ideal thermal radiator that completely absorbs all incident radiation, whatever the wavelength, the direction of incidence, or the polarization. It is also referred to as Planckian radiator. The research on and description of the behavior of thermal radiators is based on this ideal thermal radiator. Black body radiation is also the basis for the correlated color temperature of gas discharge and solid-state light sources.

The thermal radiation of a black body radiator depends solely on the temperature of the material (T) and is completely described by three laws: the law of Planck, the law of Stefan-Boltzmann, and the displacement law of Wien. The law of Planck defines the continuous spectrum of the black body radiator. The law of Stefan-Boltzmann states that the total radiant flux increases proportionally to T^4 . Wien's displacement law defines how the peak of the spectrum shifts to shorter wavelengths with an increase of T (i.e., from the infrared region of the spectrum via the red part of the visible spectrum to the blue part of the visible spectrum).

Incandescent material used for actual lamps such as tungsten filaments of incandescent and halogen lamps behaves similarly to the black body radiator but with some important differences.

Characteristics of Commercial Incandescent Lamps

The efficacy of incandescent lamps is with some 5–30 lm/W (dependent on type and wattage) low compared to that of gas discharge lamps and solid-state LED light sources with values between 50 and 200 lm/W. The average lifetime with values between 1,000 and 3,000 h is short relative to the economic lifetime of gas discharge lamps and LED lamps with values varying, with type, from some 10,000–10,000 h. The temperature of the incandescent filament is limited to some 3,200 K because at higher temperatures the material will evaporate quickly or even start to burn. The consequence is that the continuous spectrum of both the incandescent and halogen lamps has its

peak in the red part of the spectrum. The peak with the halogen type is slightly more to shorter wavelengths. The color of the light is warm white (color temperature between 2,700 and 3,200 K). Since the spectrum is continuous, the color rendering index is 100.

Cross-References

- [Carbon Arc Lamp](#)
- [Combustion Lamp](#)
- [Halogen Lamp](#)
- [Incandescent Lamp](#)

Incandescent Bulb

- [Incandescent Lamp](#)

Incandescent Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Incandescent bulb](#); [Tungsten filament lamp](#)

Definition

A lamp that produces light as a result of electrical current through a tungsten wire contained in a transparent bulb, which heats the metal to incandescence.

Incandescent Lamps

There exists a wide variety of incandescent lamps, ranging from the most common types (usually referred to as general lighting service, GLS,

lamps) to reflector lamps, colored lamps, and halogen incandescent lamps [1, 2]. Given the special operating principle and construction of halogen incandescent lamps, these will be dealt with separately under “halogen, halogen lamp.” The most important application area of normal incandescent lamps is home lighting. Because of the low efficacy and short lifetime of incandescent lamps, the more efficient and longer-life gas discharge and solid-state lamp alternatives have become increasingly more important for home lighting. For reasons of sustainability, governments in some parts of the world have taken the decision (e.g., Australia and the European Union), or are considering making the decision, to ban GLS incandescent lamps.

Working Principle

The operating principle of the incandescent lamp is simple. An electric current is passed through a thin wire of comparatively high resistance so as to heat this to incandescence. The wire is usually heated to a temperature of between 2700 and 2800 K, at which temperature it emits warm-white light. The wire is placed inside a glass bulb, which is either at a vacuum or contains an inert gas.

Materials and Construction

The main parts of an incandescent lamp are (Fig. 1):

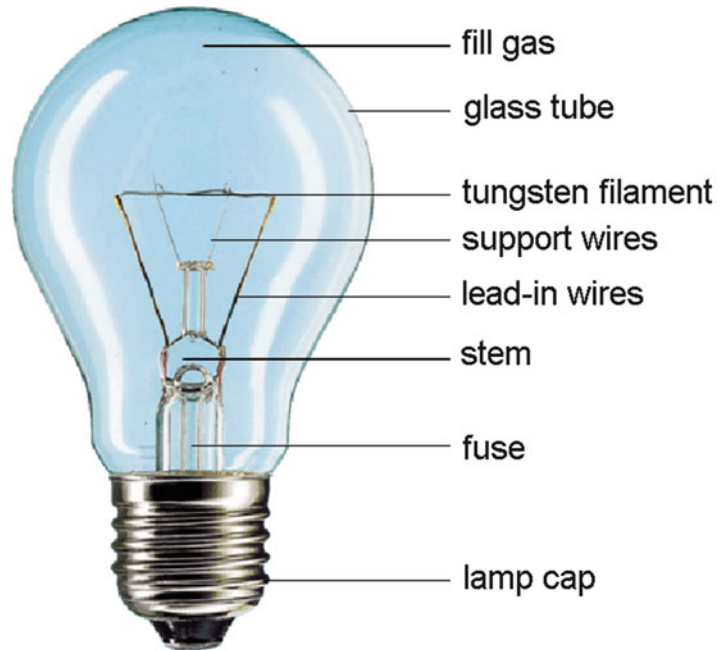
- Glass bulb
- Filament
- Filament support consisting of a glass stem, lead-in wires (also serving as a fuse), and support wires
- Bulb fill gas
- Lamp cap

Filament

With very few exceptions, filaments for incandescent lamps are made of tungsten, a metal that has the advantage of a high melting point combined with a relatively low vapor pressure, even at very high temperatures. This low vapor pressure means

Incandescent Lamp,

Fig. 1 The main parts of an incandescent lamp [3]

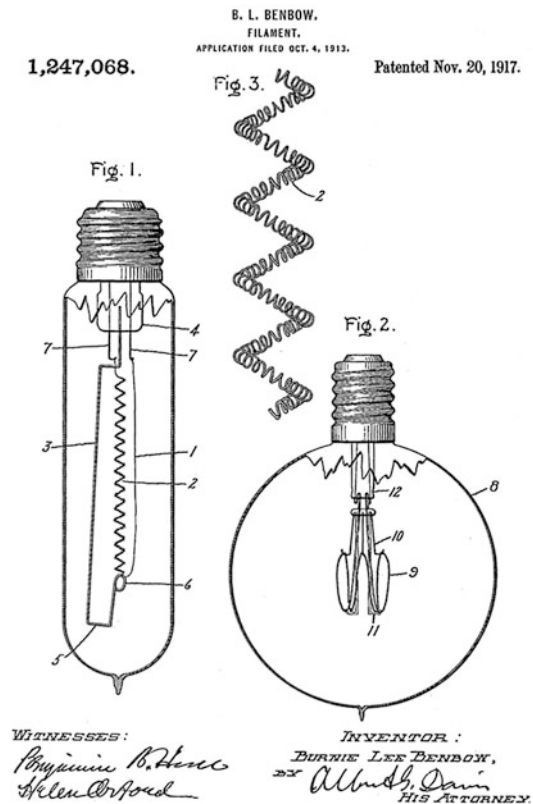


a low evaporation rate of the tungsten material and thus a relatively long life of the coil. The tungsten wire is generally wound into a single or double coil for reasons of compactness and heat conservation and concentration around the coil. The double coil (Figs. 2 and 3), which conserves and concentrates the heat better than does a single coil, was patented in 1917. It was the last fundamental improvement of incandescent lamps as far as the operating principle is concerned.

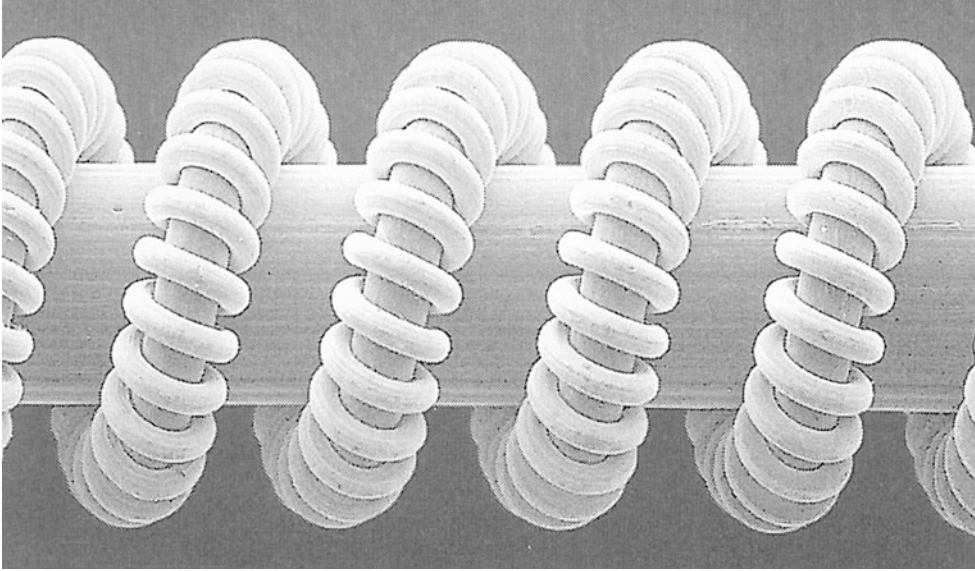
Bulb

The bulbs of GLS lamps are made of soda-lime glass, the most common and cheapest type of glass available. For lamps that must withstand high temperatures or temperature chocks, more resistant glasses are used. The bulb of an incandescent lamp can be of different sizes and may take different shapes, depending on the application. Also depending on the application, the bulb may undergo various treatments. The most commonly employed treatments are:

- **Frosting:** This is done by etching the inside of the bulb. The etching of the surface results in a satin finish and moderate diffusion of the light with hardly any reduction in transmittance.



Incandescent Lamp, Fig. 2 Patent of 1917 of the double-coiled tungsten wire



Incandescent Lamp, Fig. 3 Winding of the tungsten wire into a double coil around a core, which is removed later on in the manufacturing process [3]

- **Opalizing:** This is achieved by coating the inside of the bulb with a special powder. It provides better diffusion of the light at the cost of slightly larger light absorption.
- **Mirror coating:** Reflector lamps receive an internal mirror coating. Silver mirrors are produced by evaporation of aluminum under vacuum.

Fill Gas

Two types of incandescent lamp are available on the market; vacuum lamps and those with a fill gas. The vast majority of incandescent lamps contain a fill gas, the primary purpose of which is to reduce the evaporation rate of the filament and thus increase the lamp life. The fill gas is a mixture of argon and nitrogen. In incandescent car lamps, krypton is usually employed (it is more expensive, but results in a higher luminous efficacy).

Even small quantities of oxygen or water vapor could shorten the lifetime of the tungsten wire through corrosion. To remove even the slightest traces of these, a so-called getter is added. The getter absorbs oxygen and water.

Lamp Cap

The lamp cap connects the lamp to the power supply socket. There are two different standardized cap types: the screw cap (also called the Edison cap) and the bayonet cap. Both are employed in various sizes:

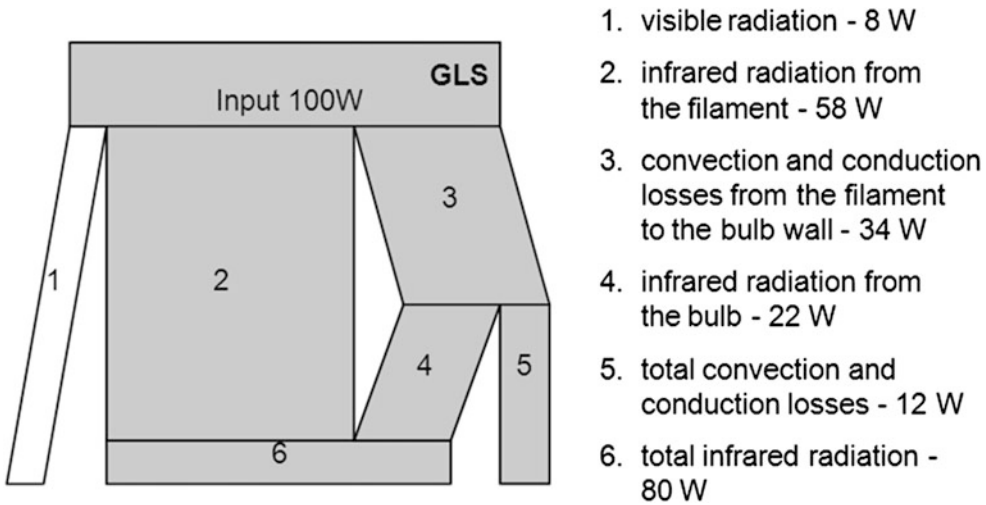
- Edison screw type: IEC designation – E10, E14, E27, E40. In the USA, names rather than IEC designations are used: Candelabra (E12), Intermediate (E17), Medium (E26), and Mogul (E39).
- Bayonet type: IEC designation B15, B22.

The number indicates the diameter of the lamp cap in millimeters.

Properties

Energy Balance

Figure 4 shows the energy balance of a 100 W GLS lamp. It shows that only approximately 8% of the input power is emitted in the form of visible radiation. The rest is lost as heat (by conduction, convection, and infrared radiation).



Incandescent Lamp, Fig. 4 Energy balance of a 100 W GLS lamp [3]

Luminous Efficacy

The luminous efficacy of a standard GLS lamp varies, depending on its wattage, between 8 and 15 lm/W. For a standard 75 W lamp, the efficacy is around 12 lm/W. The lower the wattage, the lower the efficacy.

Compared with gas discharge and solid-state light sources, the efficacy of incandescent lamps is very low indeed. For reasons of sustainability, governments in some parts of the world (e.g., the European Union) have taken the decision, or are considering making the decision, to ban GLS incandescent lamps.

Lumen-Package Range

Common types of GLS lamps are available in the range from some 100 to 1500 lumen (corresponding wattage range approx. 15–200 W). There are, of course, many special lamps with both lower and higher lumen packages: think, for example, of bicycle and torch lamps on the one hand and light-house lamps on the other.

Color Characteristics

Incandescent lamps have a continuous spectrum radiating more energy at longer wavelengths (red) than at shorter wavelengths (blue) (Fig. 5).

Their color rendering index is 100, and standard GLS lamps have a color temperature of between 2700 and 2800 K (warm-white light).

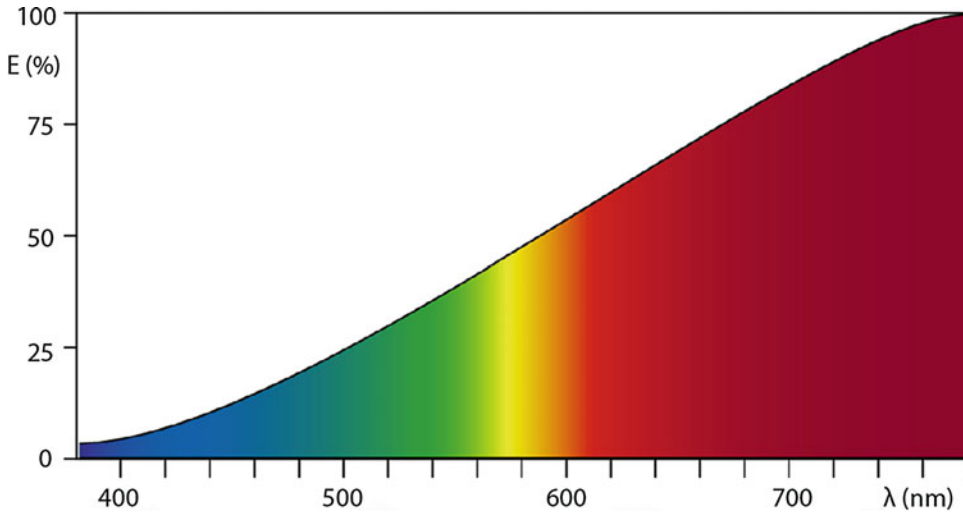
Colored opal lamps for party or festive lighting are available, as are lamps producing a special tint of “white” light (e.g., flame, terracotta, beige).

Lamp Life

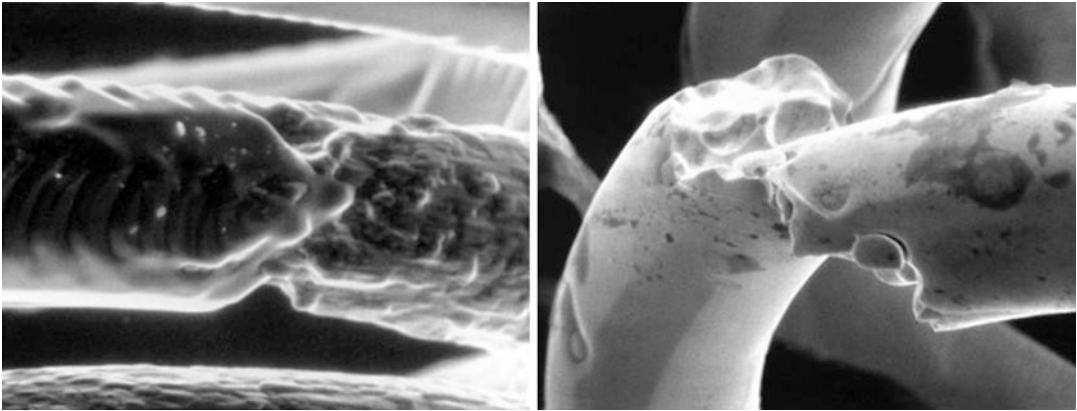
Lamp manufacturers can, in principle, balance the life of an incandescent lamp against its luminous efficacy: higher lifetime, lower efficacy. However, by international agreement, the lamp industry has standardized the average rated life of standard GLS lamp at 1000 h (750 h in the USA). Nevertheless, for special applications, lamps are available where the balance between lifetime and efficacy is different. Examples are photo and film-studio lamps and torch bulbs on the one hand with their shorter life but higher efficacy (and higher color temperature) and beacon lamps on the other with their longer life but lower efficacy.

Lamp-Lumen Depreciation

An important reason why the light output of an incandescent lamp decreases with time is that tungsten evaporates from the filament and settles on the bulb wall, blackening it. At the end of the



Incandescent Lamp, Fig. 5 Relative spectral energy distribution of an incandescent lamp [3]



Incandescent Lamp, Fig. 6 A thin spot in the filament eventually resulting in breakage of the filament [3]

rated average life (1000 h), the lumen depreciation can amount to as much as 15–25%.

Run-Up and Re-Ignition

Incandescent lamps give their full light output immediately after being switched on, and after being switched off, they re-ignite again immediately.

Switching

Frequent switching usually is not detrimental to lamp life, but when the filament has become

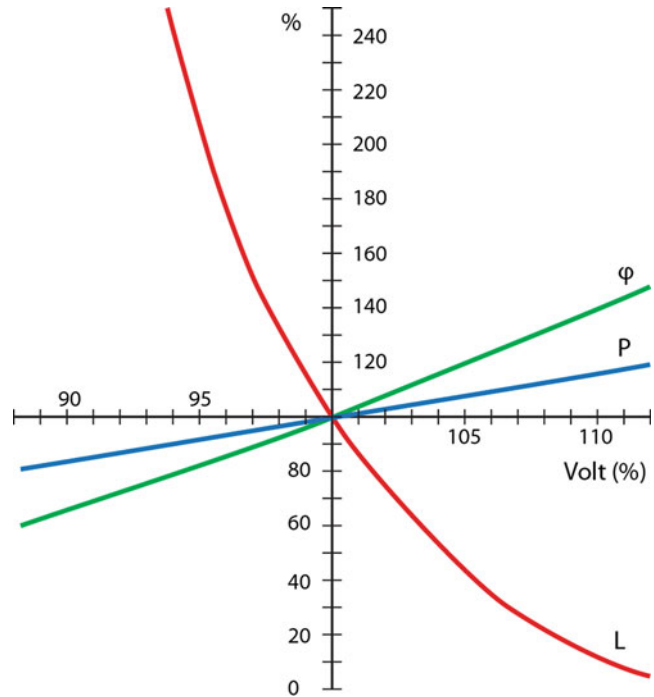
critically thin through age, the mechanical strain caused by the rapid temperature change as a result of switching will be sufficient to cause its breakdown (Fig. 6). This is the reason why incandescent lamps approaching the end of their life usually fail the moment they are switched on or off.

Dimming

Regular incandescent lamps can be dimmed without restriction. Dimming by reducing the supply voltage to the lamp with an adjustable resistance

Incandescent Lamp,

Fig. 7 Diagram indicating the effects of mains voltage variation on lifetime (L), luminous flux (ϕ), and dissipated power (P)



Incandescent Lamp, Fig. 8 Examples of general lighting service (GLS) incandescent lamp types

in series with the lamp is technically easy but does not result in energy saving. A relatively expensive adjustable transformer (potentiometer) results in energy saving, but not very much. Today the usual method to dim incandescent lamps is by thyristor dimmers that cut part of the AC current during each of the 50 (Europe) or 60 (USA) cycles.

During the time that the current is cut, no power is dissipated, so that indeed energy is saved. These systems are called phase-cutting systems. They are small enough to fit into a wall switch box or cord switch. A dimmed incandescent lamp will have a lower filament temperature, which results in longer life and lower color temperature (warm-



Incandescent Lamp, Fig. 9 Examples of refractor/reflector incandescent lamps

colored light). The lamp efficacy while dimming decreases, which means that the percentage of power saved is less than the percentage reduction in light output.

Mains Voltage Variations

An overvoltage of even a few percent results in a drastically reduced lamp life. For example, a permanent overvoltage of 5% reduces lamp life by 50% (Fig. 7).

Product Range

The product range of incandescent lamps comprises not only general lighting service lamps (in clear, frosted, opal, and colored versions) but also reflector lamps, tubular lamps, decorative lamps, beacon lamps, vehicle lamps (bicycles, cars, trains), signal lamps, and studio and theater lamps. Figures 8 and 9 show some examples of these lamps. All these lamps have blown-glass bulbs (blown from a single “blob” of glass), with the exception of the pressed-glass reflector lamp (PAR). The latter is molded in two pieces, which are sealed together – they are also sometimes called sealed-beam lamps. One advantage of this construction is that the mirror in the rear and the refractor on the front can be accurately shaped to increase the beam quality and efficiency. Another advantage is its high mechanical and thermal

strength, allowing the lamp to be used outdoors without a protecting luminaire.

Cross-References

► [Incandescence](#)

References

1. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
2. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. Illuminating Engineering Society of North America, New York (2011)
3. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)

Incoherent Scattering

► [Compton Scattering](#)

Indirect Illumination

► [Global Illumination](#)

Individual Differences

► [Environmental Influences on Color Vision](#)

Indoor Lighting

► [Interior Lighting](#)

Induction Lamp

Wout van Bommel
van Bommel Lighting Consultant,
Nuenen, The Netherlands

Synonyms

[Electrodeless low-pressure mercury lamps](#)

Definition

Lamps that produce light as a result of an electrical discharge, generated by an induction coil, in a

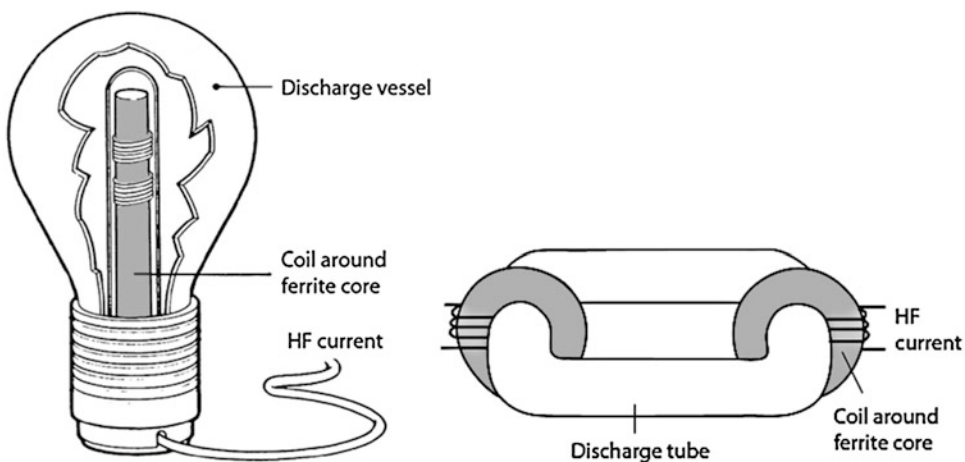
low-pressure mercury vapor that is contained in a transparent tube or vessel whose inside is coated with a fluorescent powder that converts the ultra-violet part of the emitted radiation from the discharge in visible light.

Induction Lamps

Induction lamps, like fluorescent lamps, belong to the family of low-pressure mercury gas discharge lamps. Unlike other discharge lamps, they have no electrodes, which is why they are also called “electrodeless lamps” [1, 2]. The consequence of having no electrodes is a very long economic life of around 60,000–75,000 h. This long life is also the main feature of induction lamps. They find their application in situations where lamp replacement is near impossible or very expensive.

Working Principle

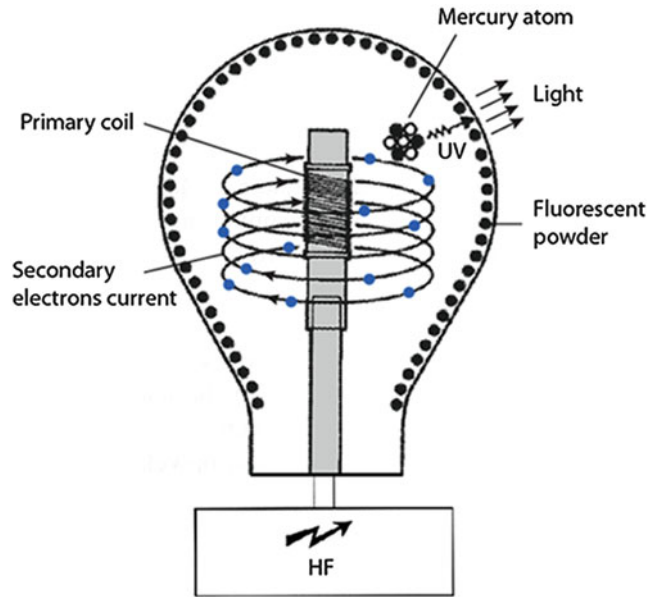
In an induction lamp, the free-running electrons needed for the gas discharge are obtained by winding an induction coil around a ferrite core placed in or around a discharge vessel (Fig. 1). The induction coil is connected to a high-frequency power source and acts as a primary winding in a transformer. In a transformer, an alternating current in a primary winding around



Induction Lamp, Fig. 1 The principal parts of an induction lamp. Left, induction coil inside the discharge vessel; right, induction coils around the outer part of discharge tube

Induction Lamp,

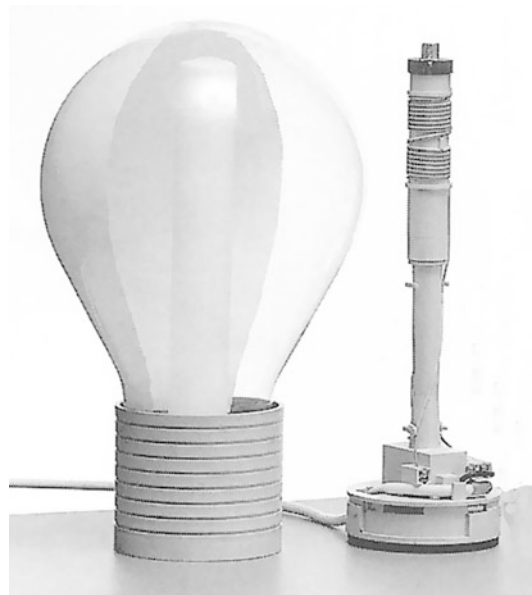
Fig. 2 Principle of an induction lamp [3]



an iron core creates an alternating magnetic field that in turn initiates an alternating current in a secondary coil wound around the primary coil. In the discharge vessel or tube, the mercury gas surrounding the ferrite core acts as the secondary coil (since mercury is a metal). The secondary current initiated in the mercury consists of free-running electrons around the ferrite core (Fig. 2). As in a normal fluorescent gas discharge, these free-running electrons ionize and excite other mercury atoms, which results in the emission of the same radiation as in a tubular fluorescent lamp. Since the vessel's interior is coated with the same fluorescent powder as in regular fluorescent lamps, the same type of light is obtained. The high frequency of 2.65 MHz is generated by an electronic circuit.

Materials and Construction

The discharge vessel or tube is made of the same type of glass as regular fluorescent tubes and has, in the case of the internal coil, a cylindrical glass cavity in which the ferrite core with coil (also called "antenna") is positioned (Fig. 3). Special precautions are taken in the design of the lamp to limit excessive heat buildup around the antenna. The high-frequency power generator is connected to the antenna with a coax cable. The cable forms



Induction Lamp, Fig. 3 Photograph of induction lamp vessel and antenna. This lamp is only partly covered with fluorescent powder in order to show the cylindrical cavity in which the coil is positioned [3]

part of the electronic circuit and is, therefore, an integral part of the system (Fig. 4). The version with the coil with iron core around the outside of the discharge tube has a different, planar shape than the pear-shaped version.



Induction Lamp, Fig. 4 Pear-shaped induction lamp with an internal coil (left) and a plane-shaped lamp with external coils

Properties

Energy Balance

Around 17% of the input power of an induction lamp is radiated as visible light. The remaining part is lost as heat in the power generator, in the antenna, and in the discharge.

System Luminous Efficacy

The luminous efficacy of induction lamps is smaller than that of regular fluorescent lamps. It varies, depending on the wattage, between roughly 65 and 75 lm/W.

Lumen-Package Range

Induction lamps are available in a lumen-package range from some 3500 to 12,000 lm (corresponding wattage range 55–165 W).

Color Characteristics

The fluorescent coating is of the same composition as that in regular fluorescent tubes, so the color characteristics are also the same. See chapter ► [“Tubular and Compact Fluorescent Lamp.”](#)

Lamp Life

The lamp life of most induction lamps is extremely long. Based on a mortality rate of 20%, induction lamps have a lifetime of between

60,000 and 75,000 h. This is almost 7 years of continuous operation, day and night.

Lamp Price

Induction lamps are expensive. The high lamp price has to be balanced against the extremely long lifetime of these lamps. This means that where lamp replacement is difficult, very expensive, or even impossible, the balance will certainly be in favor of the longer life of these lamps.

Lamp-Lumen Depreciation

Lumen depreciation in induction lamps is determined by a decrease in the activity of the fluorescent powder. At 60,000 h the depreciation is around 25%, which is why this lifetime figure is usually quoted. The actual life of induction lamps is usually much longer.

Run-Up and Reignition

A high-voltage ignition pulse produced by the HF generator ignites the lamp within 5 s, after which it emits its full light output within 1 min. Thanks to the high-voltage ignition pulse, hot reignition of the lamp is no problem.

Dimming

Most standard versions of induction lamps are not dimmable. However, some special versions, which are dimmable to 50%, are available.

Ambient-Temperature Sensitivity

Amalgam is used instead of pure mercury to keep the sensitivity to ambient temperatures within practical limits. Lamp position sometimes plays a role in this as well. It is therefore important to follow the recommendations from the lamp manufacturer with respect to the luminaire design.

Product Range

As has been shown, induction lamps are available in versions with an internal and external antenna. The corresponding shapes are pear and plane-shaped. They are generally available in color-rendering types 800 and 900, with color temperatures of 3000 and 4000 K.

Cross-References

- [Phosphors and Fluorescent Powders](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Coaton, J.R., Marsden, A.M.: *Lamps and Lighting*, 4th edn. Arnold, London (1997)
2. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *IES Handbook*, 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)
3. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course Book. Philips Lighting, Eindhoven (2012)

Definition

A response to color in infancy that indicates that infants can divide the continuum of color into discrete groups.

Key Principles and Concepts

Although humans can discriminate millions of colors, language typically refers to color using a limited number of discrete categories (e.g., *red*, *green*, *blue*). Color categories are not only present in language (color terms), they are also present in “thought.” For example, at least under some circumstances, adults’ perceptual or cognitive judgments of color (e.g., color memory, search, or similarity judgments) are affected by whether colors come from the same or different linguistic color categories [see Hanley]. There has been considerable debate on the origin of color categories and on the extent to which color categories are arbitrarily constructed through language (see ref. [1] for review). To establish whether there is a nonlinguistic route to color categorization, prelinguistic infants’ response to color has been investigated. This has provided converging evidence that infants’ perceptual or cognitive judgments of color can also be affected by whether colors come from the same or different adult color categories. This suggests that color categories are present even in infancy and that even infants, on some level, can parse the continuum of color into discrete groups.

Methods and Examples

Investigations of adults’ categorical response to color have used a range of experimental tasks that provide behavioral or electrophysiological measures of color memory, color search, or chromatic change detection. In these studies, chromatic differences for same- and different-category color pairs are equated, and an influence of the categorical relationship of colors on task performance is assessed. An equivalent approach has been taken to investigate infant color categories. Many of the

Infant Color Categories

Anna Franklin

Psychology, University of Surrey, Surrey, UK

Synonyms

[Prelinguistic color categorization/categorical perception](#)

studies of infant color categories have used the well-established method of *habituation* to assess the influence of categories on infants' response to chromatic novelty. This method involves the repeated presentation of a stimulus (or set of stimuli) until infants' looking time at the stimulus declines (habituation). Then, during a test phase, novel and original stimuli are presented and looking time is recorded. Greater looking to the novel stimulus relative to the original stimulus during the test phase (novelty preference) indicates that the infant is, on some level, treating novel and original stimuli as different. A lack of longer looking at the novel stimulus relative to the original indicates that the infant is, on some level, treating novel and original stimuli as equivalent. The degree of novelty preference can therefore be compared when novel and original stimuli from the same or different categories. Greater different-than same-category novelty preference when chromatic differences are equated would indicate an influence of categories on infants' perception or cognition. In addition, categorization can be broadly defined as the treatment of discriminably different stimuli as equivalent (e.g., [2]). Therefore, a lack of novelty preference when infants can discriminate between novel and familiar stimuli becomes a clear marker of categorization. This method has provided overwhelming evidence for infant categories, for a wide range of categories such as phonemes, facial expressions, animals, spatial locations, and orientation (e.g., see [2]).

For color, multiple studies have found a categorical response in infants at 4 months using the habituation method (e.g., [3–6]). Bornstein et al. [3] tested infants at 4 months using monochromatic lights, with the difference between the original and novel stimulus equated in wavelength. Infants responded to a change in stimulus that crossed blue-green, yellow-green, and yellow-red category boundaries, but not to a change in stimulus within these categories. There was one exception to this: when infants were habituated to a red wavelength, their interest in the original red stimulus continued at test, most likely due to the salience of red in infancy. Apart from this, infants' recognition of novelty in hue was completely categorical, and Bornstein et al. concluded that

infants categorize blue, green, red and yellow hues at 4 months. Two further studies also provide evidence of a categorical response to color in infants, finding that 4-month-olds fail to recognize the novelty of a color when original and novel colors are from the same blue or green category [4] or from the same red or blue categories after a 5 min delay between habituation and test [5]. In addition, a fourth study indicates that 4-month-old infant's response to chromatic novelty is also categorical across blue-green, red-pink, and blue-purple category boundaries, when same- and different-category colors are equated in color metrics such as the Munsell color system and CIE ($L^*u^*v^*$, 1976) color space [6].

Therefore, a categorical response to color in infancy has now been found in multiple habituation studies, for over 25 different stimulus pairs taken from three different color metrics, and with different versions of the habituation task. In all of these studies, the same-category chromatic differences are known to be discriminable to infants at 4 months. Infants at 4 months are able to discriminate same-category chromatic differences of an equivalent size when tasks do not involve color memory (e.g., color search: [7, 8]), and the same-category chromatic differences were all much greater than chromatic discrimination thresholds at 4 months. In fact, some of the same-category stimulus separations in Bornstein et al. investigation were as much as 20 times larger than the estimated just-noticeable difference for a 4-month-old (see [6]). Even when a luminance difference is intentionally added to maximize the difference between novel and original stimuli [6], 4-month-old infants still fail to respond to the novelty of a color from the same category as the original. Infants' pattern of responding to color on habituation tasks therefore fits the classic definition of categorization that "discriminably different stimuli are treated as equivalent" (e.g., [2]). These studies provide a parallel to adult studies which have used an X-AB task. The X-AB task requires participants to encode an original color and then distinguish it from a novel foil after a delay. Adults are poorer at identifying the original color when the original and foil are from the same category than when they are from different

categories. Infants' pattern of response on habituation tasks is equivalent to this, as they are also poorer at distinguishing between original and novel same- than different-category colors at test. In fact, infants are even more categorical in their response than adults as, at least under some conditions, they fail to distinguish between the original and novel same-category colors at all.

Another method which has been used to investigate both adult and infant color categories is the event-related potential (ERP) technique. This involves measuring electrical activity from the scalp, which is then time locked to an event such as stimulus onset. Event-related potentials have been measured on a visual oddball task which is commonly used in ERP studies of both adult and infant categorization. The visual oddball task involves the frequent presentation of one stimulus (the standard) interspersed with the infrequent presentation of other stimuli (the deviants), and ERP components for standard and deviant stimuli are compared for evidence of change detection. For adults, the categorical relationship of standard and deviant colors affects the latency and amplitude of several ERP components elicited in response to the stimuli. For example, different-category deviants elicit ERP components that peak earlier or are of greater amplitude than same-category deviants, indicating quicker and greater detection of different- than same-category chromatic change (see [9] for review). Category effects in ERPs on a visual oddball task have also been found in infancy [9]. Even at 7-months, the categorical relationship of standard and deviant colors affected infants' ERP components. Category effects were found in ERP components thought to index recognition memory (slow waves) and attentional allocation (negative central). For both, the different-category deviant elicited an ERP component that was more negative than that elicited by the standard color, yet the same-category deviant did not. These findings indicate that, within the context of the frequent repetition of a color, an infrequent different-category color will elicit greater attentional allocation and will be seen as more novel than the frequent color, whereas a same-category infrequent color will not. As in the habituation studies,

the same-category colors were discriminable to infants at 7 months, and therefore the equivalence of the same-category deviant and standard in ERP components of attention and memory indicates electrophysiological markers of infant color categorization. Due to the time course of the negative central ERP component, Clifford et al. study also suggests that infants register the categorical status of a color from as early as 250 ms poststimulus onset. These findings are in line with other ERP studies of categorical responding in infancy, such as phoneme categorization and category learning of cats and dogs (see [9]).

Chromatic search tasks have also been used to investigate infant color categories. In adult studies that have used these tasks, participants are required to search for a colored target that is either from the same or a different category to the colored distracters or colored background. When same- and different-category colors are equated in CIE ($L^*u^*v^*$, 1976) or Munsell color metrics, at least under some conditions, adults are faster or more accurate at searching for targets amongst different- than same-category distracters or backgrounds. The influence of color categories on infants' color search has been investigated by measuring 4–6-month-old eye movements to colored targets presented on colored backgrounds. Infants are significantly faster at fixating a colored target [7] and at initiating an eye movement to a colored target [8] when presented on a different- than same-category colored background. In these infant studies, the chromatic difference between the target and background was equated for same- and different-category color pairs in CIE $L^*u^*v^*$ color space yet are also closely equated when stimuli are converted to the MacLeod-Boynton cone excitation color space. Franklin et al. [8] found this effect of categories on infants' target detection when the target was presented to the left visual field, but not to the right visual field, potentially suggesting that there is a right hemisphere lateralization of categorical perception of color in infancy. This right hemisphere lateralization contrasts with reports of left hemisphere lateralization in the adults' category effect (see [1]), and it appears as though a right-to-left hemisphere switch occurs around the time of color term

acquisition [10]. However, these effects have so far only been tested for in infants across the blue-green category boundary, and recent studies also report a lack of hemispheric lateralization for the effect in adults under some conditions (e.g., [11]). Further research is therefore needed to establish the nature of these effects in both infants and adults and the conditions for the pattern of hemispheric lateralization.

Negative Findings

There is only one published study which fails to find a categorical novelty response in infants [12], and that study tested for a categorical response across the red-orange category boundary using a task which determined whether a novel color “popped out” amongst six instances of a color to which infants had been familiarized. This could indicate that infants do not respond categorically across the red-orange boundary or that the novelty response is not categorical under conditions of “pop out.” There may well be conditions under which infants’ responding is not categorical, and as for adults, the relative strength of the categorical and physical code may depend on cognitive demands. However, it has been argued that an unintended category boundary for one of the same-category stimulus pairs, and the use of an inappropriate illuminant, renders the findings of Gerhardstein et al. unsafe [13]. A second unpublished study has been cited as evidence for a lack of replication of an infant category effect across the blue-green boundary. However, the habituation task that was used failed to detect any response to color in the infants at all, most likely due to particular parameters of the habituation task (e.g., a change from central to peripheral stimulus presentation at test could have detracted attention from the change in color). Therefore, as there was no response to color at all, a categorical effect could not be assessed (see [14] for further discussion). These negative findings, which should be considered as unsafe, are outweighed by the converging behavioral and electrophysiological evidence for a categorical response to

color in infancy. The finding of infant color categories has now been replicated across multiple color category boundaries, using multiple tasks and color metrics, and by several independent research teams.

Alternative Explanations

As outlined above, there is converging evidence that, at least under some conditions, infants’ response to chromatic novelty or recognition memory can be completely categorical. Potential alternative explanations for infants’ response to color in these studies have been considered and ruled out. First, a priori hue preferences are unable to account for the findings as studies have controlled for these (see [14]). For example, same-category stimulus pairs have been sampled from both sides of the category boundary (e.g., [3, 6]), a priori preferences have been recorded and shown not to account for infants’ response (e.g., [6]), and control tasks have also been included (e.g., [9]). Second, potential inequalities in the color metric are unable to account for infants’ response to color. As for studies of adult color categories, there is an issue about which color metric is appropriate to equate same- and different-category colors: using a physical metric (e.g., wavelength), a perceptual metric such as Munsell or CIE ($L^*u^*v^*$), or a cone-opponent color space (e.g., MacLeod-Boynton color space). Importantly, infant color categories appear to transcend this metric issue. Infants’ pattern of response is identical across studies which vary in the color metric used. Additionally, at least under some conditions, infants’ novelty response to color appears completely invariant to the color metric, as distances in any of these metrics do not predict the degree of infants’ chromatic novelty preference or change detection. Infants treat same- but not different-category chromatic differences as equivalent on habituation or visual oddball tasks, irrespective of the size of the same-category chromatic difference. Only the categorical relationship between stimuli can account for this pattern of response.

Theoretical Questions and Future Directions

The converging evidence for infant color categories strongly suggests that there is a nonlinguistic route to parsing the continuum of color into discrete groups. However, there is also convincing evidence that the influence of color categories on adults' cognition and perception is dependent on language and that different languages also vary in how they divide up the color spectrum [see Roberson]. This raises two sets of challenging questions for future research to consider and address. First, what is the relationship between prelinguistic and linguistic color categories? Why do some languages not have color terms for categories that infants have? For example, many of the world's languages do not have separate terms for blue and green, yet infants appear to form separate categories for these. The answer to this question may be in the analysis of the World Color Survey [see TBA]. Multiple analyses of the World Color Survey strongly suggest that, despite variation in color lexicons, there is also striking commonality in how the world's languages divide up the color spectrum. For example, the centers of color categories across languages tend to cluster around particular points in the color space, with distinct clusters for "blue" and "green." It has therefore been suggested by several groups that there are perceptual constraints on how categories form in color lexicons (see [1]). Therefore, although not all languages have separate terms for blue and green, there appears to be a common pressure in the world's color lexicons for separate categories of blue and green to form. One possibility is that this commonality arises because infant color categories partially constrain the world's color lexicons, while local variation arises due to the additional influence of culture and communication [1].

Alternatively, infant color categories and those in language could operate independently from each other but could have similar organizing principles. This raises the question of what happens when prelinguistic and linguistic color categories converge when color terms are learnt. What are the mechanisms of the switch from a nonlinguistic

to a language dominant division of the color spectrum? How do prelinguistic color categories affect color term acquisition? The presence of infant color categories strongly suggests that any difficulty that children have in learning color words is unlikely to be due to a lack of category structure but rather provides support for the argument that there are general attentional, conceptual, and linguistic constraints that make learning color words difficult relative to nouns, but not other adjectives (see [15]). However, it is theoretically possible that color term learning may be more difficult for lexicons which map less well onto the prelinguistic category structure that is there in infancy.

Second, what are the organizing principles of infant color categories? Where do infants get their category structure from? Several theories have been proposed to identify the organizing principles of color categories. For example, it has been proposed that color categories derive from (i) the clustering of colors taken from surfaces in the natural chromatic environment; (ii) reflection properties, with some colors (those that correspond to the unique hues) reflecting incoming light in a much simpler way than others; or (iii) an irregularly shaped perceptual color space combined with basic principles of categorization (see [1]). None of these organizing principles rely on language, and all are contenders for explaining how infants are able to categorize color even before color terms have been learnt. Importantly, color categories may not be inbuilt into our visual system but rather could be a cognitive response to inequalities in our visual system's processing of color. Some hues could be more salient than others, for example, due to having greater maximum levels of saturation or different reflection properties. This could lead to some colors being more informative than others, and inequalities such as these could provide structure for cognition when categorizing. The actual categorical response to color in infants could therefore come from an interaction of perceptual inequalities in color and a general cognitive strategy to categorize. It is well known that categorization is a core part of infant cognition and is essential in helping infants to navigate an uncertain world (e.g., [2]). If

color categories do result in part from general cognitive strategies, differences between humans and animals in the extent of the categorical response to color could arise due to differences in cognition, even if color vision is similar (e.g., baboons; see Weber and Changizi).

These questions on how infant color categories relate to those in language, and on the organizing principles and underlying mechanisms of infant color categories, provide clear direction for future research. Investigations which empirically address such questions will resonate beyond color and will have implications for fundamental questions on how categories are formed, how categories in language and thought interact, and how categories are expressed in the brain.

Cross-References

- [Ancient Color Terminology](#)
- [Color Categorical Perception](#)
- [World Color Survey](#)

Acknowledgments The writing of this entry was supported by a European Research Council Grant to AF (Project CATEGORIES, ref: 283605).

References

1. Regier, T., Kay, P.: Language thought and color: Whorf was half right. *Trends Cogn. Sci.* **13**, 439–446 (2009)
2. Mareschal, D., Quinn, P.C.: Categorization in infancy. *Trends Cogn. Sci.* **5**, 443–450 (2001)
3. Bornstein, M., Kessen, W., Weiskopf, S.: Color vision and hue categorisation in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 115–129 (1976)
4. Catherwood, D., Crassini, B., Freiberg, K.: The nature of infant memory for hue. *Br. J. Dev. Psychol.* **5**, 385–394 (1987)
5. Catherwood, D., Crassini, B., Freiberg, K.: The course of infant memory for hue. *Aust. J. Psychol.* **42**, 277–285 (1990)
6. Franklin, A., Davies, I.R.L.: New evidence for infant colour categories. *Br. J. Dev. Psychol.* **22**, 349–377 (2004)
7. Franklin, A., Pilling, M., Davies, I.R.L.: The nature of infant color categorisation: evidence from eye-movements on a target detection task. *J. Exp. Psychol.* **91**, 227–248 (2005)
8. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T.: Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3221–3225 (2008)
9. Clifford, A., Franklin, A., Davies, I.R.L., Holmes, A.: Electrophysiological markers of color categories in the infant brain. *Brain Cogn.* **71**, 165–172 (2009)
10. Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T., Davies, I.R.L.: Lateralization of categorical perception of color changes with color term acquisition. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 18221–18225 (2008)
11. Witzel, C., Gegenfurtner, K.: Is there a lateralized category effect for color? *J. Vis.* **11**, 1–25 (2011)
12. Gerhardstein, P., Renner, P., Rovee-Collier, C.: The roles of perceptual and categorical similarity in colour pop-out in infants. *Br. J. Dev. Psychol.* **17**, 403–420 (1999)
13. Davies, I.R.L., Franklin, A.: Categorical similarity may affect colour pop-out in infants after-all. *Br. J. Dev. Psychol.* **20**, 185–203 (2002)
14. Franklin, A.: Pre-linguistic categorical perception of color cannot be explained by color preference: response to Roberson and Hanley. *Trends Cogn. Sci.* **13**, 501–502 (2009)
15. Franklin, A.: Constraints on children's color term acquisition. *J. Exp. Psychol.* **94**, 322–327 (2006)

Influences Between Light and Color

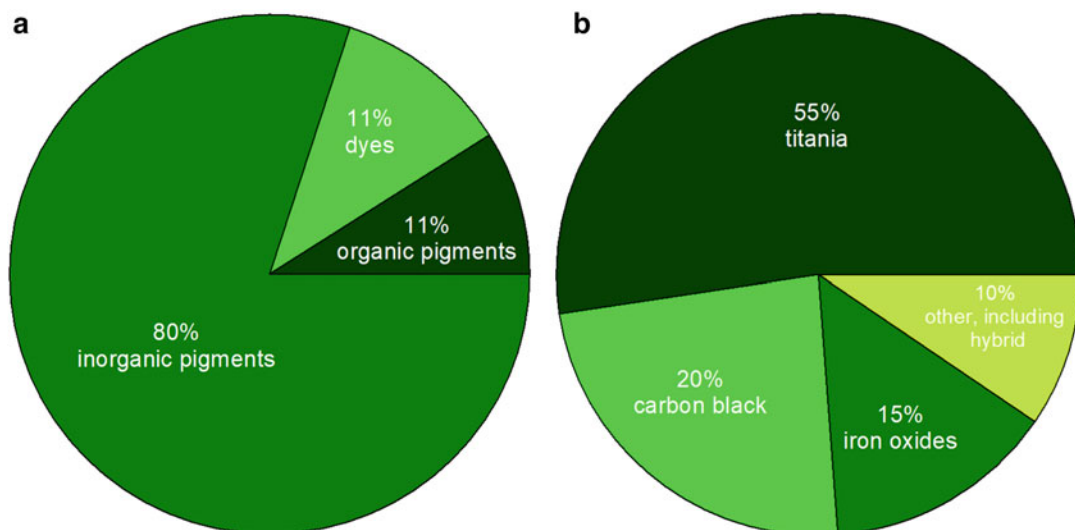
- [Interactions of Color and Light](#)

Inorganic, Hybrid, and Functional Pigments

Teofil Jesionowski and Filip Ciesielczyk
Faculty of Chemical Technology, Institute of
Chemical Technology and Engineering, Poznan
University of Technology, Poznan, Poland

Synonyms

[Biopolymers](#); [Catalysts](#); [Coatings](#); [Dyes](#); [Paints](#)



Inorganic, Hybrid, and Functional Pigments, Fig. 1 Structure of pigments consumption (a) with respect to pigment type and (b) with respect to specific pigments

Definition

Pigments, which have been widely used since prehistoric times, include naturally occurring substances prepared from minerals or their combustion products as well as synthetic compounds produced from appropriate raw materials and also hybrid pigment types derived from organic and/or natural dyes and selected mineral supports including metal oxides, biopolymers, and others. Pigments are insoluble in the surrounding medium and their optical effect arises from selective light absorption. They are classified according to their color, origin, method of production, and the character of the pigment material. Hybrid pigment technology is directly related to the development of new synthesis methods, preparation of functional and also environmentally friendly colorants.

Introduction

This entry presents the main aspects of classification and production technology, broadly

understood, for the most common inorganic pigments. It explains the most important physicochemical and application properties of a wide range of pigments, including white pigments in particular. Also discussed are issues relating to special pigments, including the new category of hybrid or functional pigments.

A clear majority of inorganic pigments are fully suitable for use in products satisfying high levels of consumer demand. Inorganic pigments are commonly used for the production of paints and lacquers, plastics, inks, construction materials, paper, glass, and ceramic products. Total consumption of pigments is presented in Fig. 1, and clearly confirms the greater importance of inorganic pigments with respect to organic ones [1].

The significance of inorganic pigments, relative to organic pigments, is consistently increasing. Inorganic pigments are more chemically inert, insoluble, and resistant to temperature and extreme pH values than organic ones. Characteristic features of inorganic pigments include excellent opacity, good color saturation, and stability. They make it possible to obtain durable and high-quality coatings.

General Information

Pigments are colored materials, which in refined state are also used for the dyeing of synthetic fabrics, rubber, and paper, and which impart color as a result of wavelength-selective light reflection, absorption, or interference. They improve the aesthetic value of certain products, and also have some practical functions (such as protecting metal objects against corrosion by painting, e.g., bridges, ships, cars, trains, and steel constructions).

Pigments are insoluble in oil, water, organic solvents, and resins. They can be classified according to color, origin, method of production, and the character of the pigment material. They are divided according to origin into inorganic and organic, and according to method of production into synthetic and natural. Inorganic and organic pigments differ in many features, such as light and weather resistance, high opacity, color strength, and ease of dispersion. Inorganic pigments usually offer better resistance to light and atmospheric conditions, higher dispersion, and opacity [2, 3].

There are pigments which, apart from coloration, also display luminescence (luminescent pigments) or change color as a function of temperature (thermosensitive pigments). These are used for the production of luminescent or thermosensitive paints.

Pigments can include different components; according to the character of the material from which they are made, they are divided into:

- Homogeneous – made of pigment particles of the same type, e.g., oxides of color metals
- Mixed or hybrid pigments – made by mechanical mixing (vigorous mixing of the dyeing substance with an extender) or by chemical synthesis (organic dyes or pigments grafted onto mineral supports or biopolymer-based supports)

Classification of Inorganic Pigments

Pigments can be classified on the basis of their chemical structure, optical properties, or technological properties.

A detailed classification of inorganic pigments according to their chemical structure and color subgroups is illustrated in Fig. 2.

Properties

Pigments are mainly used for coloration and for endowing specific properties to paint coatings. They are important components of lacquer and paint coatings, influencing their final mechanical properties and resistance; they affect the paints' rheology, stability, and some properties related to application. Those physicochemical properties of pigments chiefly influence the application and technological conditions of their commercial use [3–5].

From the application point of view, the most interesting are functional properties of pigments for specific target demands, including the following:

- Optical properties – color, tinctorial strength, opacity
- Physical – solid state properties – crystalline structure, grain morphology, character of surface, light refraction index, and defined structural properties
- Resistance and protective properties

All of these properties are to some degree determined by chemical structure [3].

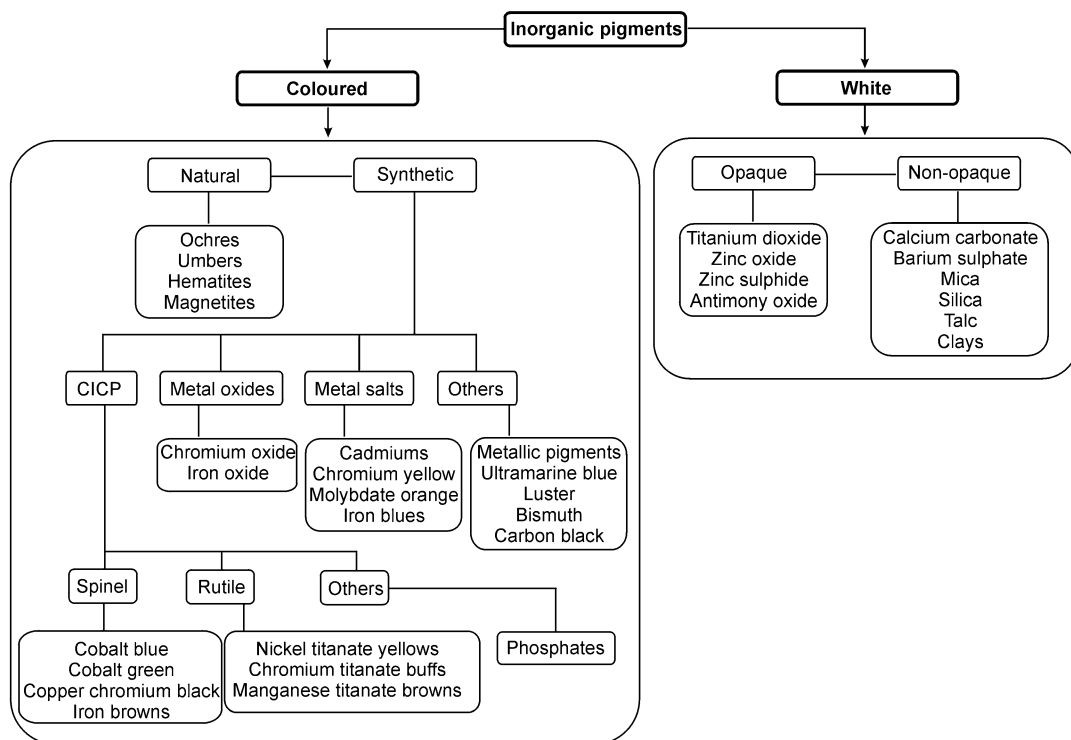
The pigment color, shade, opacity, resistance to chemicals as well as viscosity depend mainly on the chemical and crystalline structure, light refraction index, size and shape of particles, and character of the particle surface. The crystalline structure determines such properties as dispersion, rheology, solubility, color strength, and shade.

Exemplary morphological structures of selected inorganic pigments and extenders are illustrated in Fig. 3.

The fundamental properties of inorganic pigments are the following.

- *Color and intensity of coloration*

The color depends on chemical composition and degree of purity, as they define the ability to absorb certain wavelengths and to reflect the complementary color. The tinctorial strength or



Inorganic, Hybrid, and Functional Pigments, Fig. 2 Inorganic pigments classification

color strength is defined as the pigment's ability to change the original color of a given material. This is related to the pigment's introduction to a given material and depends on the degree of refinement, purity, and homogeneity of chemical composition.

- *Opacity*

The opacity (hiding power) depends on the size and structure of pigment particles and on the difference in the light refraction index of the pigment and the binding agent. The covering ability is better when the pigment has small particles and then difference in the light refraction index is significant.

- *Shape and size of pigment particles*

The size of pigment particles is regulated to obtain the optimum visible light scattering. The size of particles has an impact on the stability, coating resistance, and its covering power. The pigment particle size can vary from 1 to 10 μm , and particles of the smallest

diameters ensure higher density and protective properties.

- *Resistance to UV radiation*

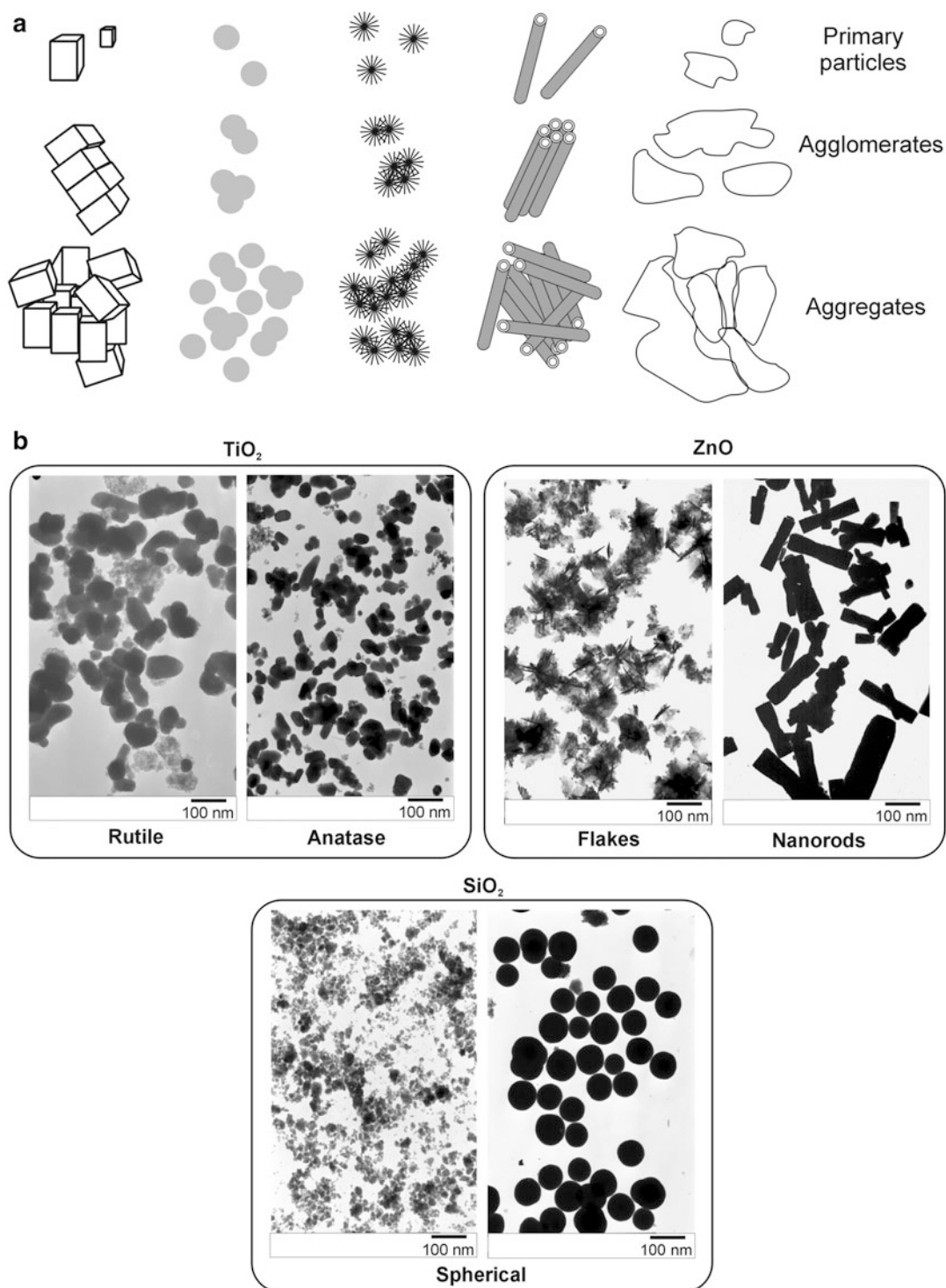
Pigments should display the highest possible resistance to light. A change in color, usually brightening or darkening under exposure to visible light or other radiation, is mainly caused by changes in the dispersion of the pigment in the binding agent and changes in that agent.

- *Thermal resistance*

Thermal resistance is of particular importance when paint coatings are made on hot substrates or when dyeing plastic products are subjected to hot processing. The pigments used for such purposes should be resistant to high temperatures.

- *Resistance to chemical agents*

Of particular importance is the pigment's resistance to alkaline and acid agents. A pigment's resistance to the binding agent,



Inorganic, Hybrid, and Functional Pigments, Fig. 3 (a) Particles morphological structure and **(b)** TEM images of commonly known inorganic white

pigments and extenders. (Created based on [3] with permission, Copyright 2013: Vincentz Verlag Publisher)

substrate, and exploitation conditions should also be analyzed. When pigments are mixed their interaction must also be established.

- *Optical properties*

Optical properties of pigments are described by the color and light absorption coefficient. The color is determined by the adsorbed or reflected wavelengths of light.

- *Wettability by binding agent*

Wettability by the binding agent depends on the degree of the pigment particles' refinement, the pigment structure, and the wetting properties of the binding agent. For oil products it is described by the so-called oil absorption number.

- *Toxicity*

Mineral pigments often contain heavy metals, but their presence is not a threat to human health or the natural environment. In the process of production of such pigments the heavy metals are so strongly bonded in the pigment structure that they are not released into the environment. In many countries the content of heavy metals in pigments for use in special products, such as those that are in contact with food, toys for children and cosmetics, is controlled by special regulations specifying admissible heavy metal levels.

Most of the aforementioned physicochemical properties of pigments determine their potential uses. Among the most common types displaying these physicochemical properties are white and black pigments, and more rarely brown, yellow, red, blue, and green pigments.

White pigments – the optical effect of these is achieved by nonselective light scattering; technologically the most important examples are titanium dioxide, zinc white, zinc sulfide, and lithopone.

Titanium dioxide (C.I. Pigment White 6; TiO_2) occurs in three crystalline varieties: brookite (not used as a pigment), anatase, and rutile. Irrespective of the technology of titanium white production, the pigments are characterized by high purity and defined crystallites and particle sizes. To improve the functional properties of the

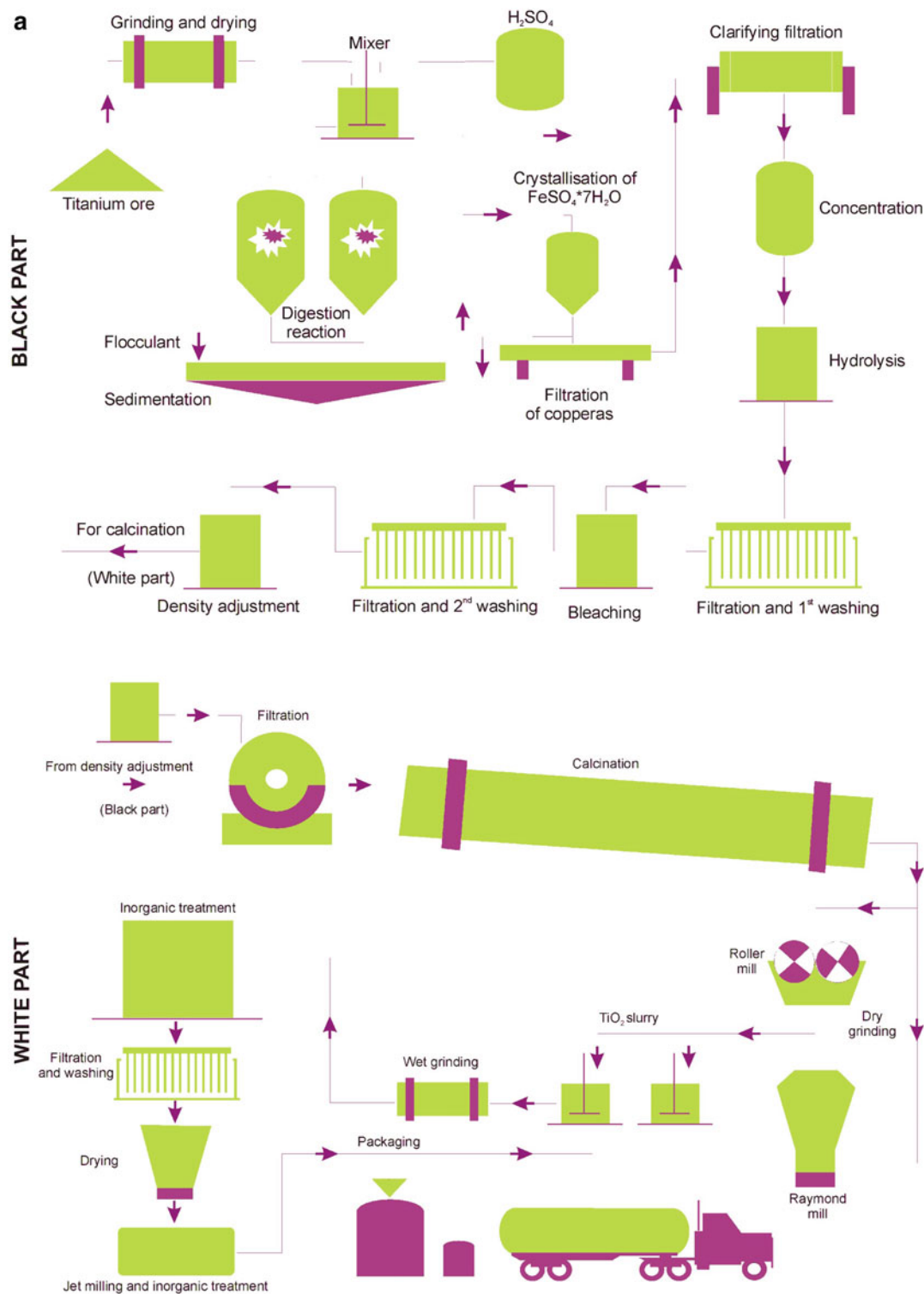
pigment in a coating, the titanium white surface is modified with different chemical compounds. The attractive features of titanium white pigment include high light refraction index, chemical and physical stability, possibility of regulation of particle sizes, and possibility of surface modification. The crystallites and particle sizes are the main parameters determining the paint properties. The optimum size of TiO_2 crystallites ensuring maximum light scattering, and hence high opacity of paint coatings, is close to $0.23\ \mu\text{m}$. The physical and chemical resistance of pigments contributes to the resistance of a paint coating. The paint coat usually undergo degradation under the influence of ultraviolet radiation, so the addition of substances absorbing UV radiation increases the paint coat's lifetime. Moreover, titanium dioxide has the highest refractive index (2.55 for anatase and 2.70 for rutile) among all inorganic pigments and commonly known extenders. Because of the universal properties of this pigment, it is produced by industrial methods on a wide scale [6, 7].

Production of TiO_2

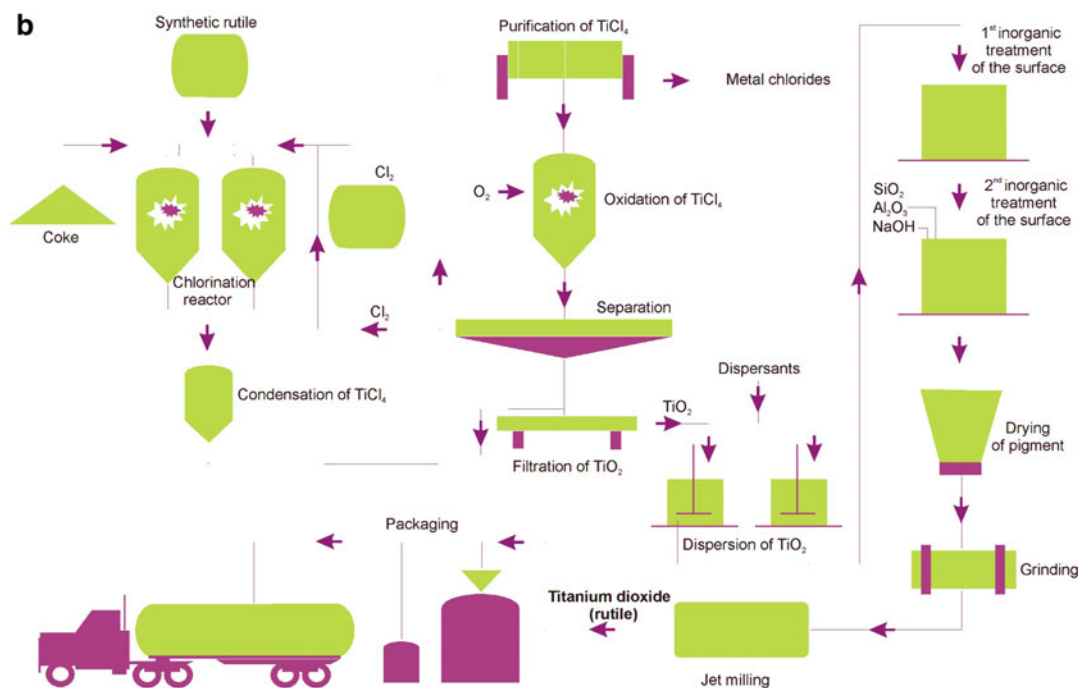
Titanium white pigments are produced on an industrial scale by two methods: sulfate and chloride.

- Sulfate process – TiO_2 is precipitated from a solution of ilmenite ore by concentrated sulfuric acid; both rutile and anatase can be obtained.
- Chloride process – TiO_2 is obtained as a result of oxidation of titanium tetrachloride (TiCl_4) synthesized by reduction and chlorination of ilmenite ore or rutile sources; only rutile is obtained.

The chloride process is more often used. It is estimated that over half of the titanium white produced worldwide is obtained by this method. Although this process needs higher quality and preliminarily enriched ore and more complex technology, it produces less waste (thus being less burdensome to the environment) and is cheaper than the sulfate method.



Inorganic, Hybrid, and Functional Pigments, Fig. 4 Production of TiO_2 by the (a) sulfate method and (b) chloride method



Inorganic, Hybrid, and Functional Pigments, Fig. 4 (continued)

Sulfate Method

The sulfate method of titanium white production is complex. It is divided into two main technological parts, called the black part (alluding to the color of the ore) and the white part (alluding to the color of white pigment). Each block includes several unit operations.

The idea is to extract TiO_2 from raw titanium ore at elevated temperature by means of concentrated sulfuric acid (85–95%). The most often used titanium ores are ilmenite (FeTiO_3) containing 45–65% TiO_2 , and titanium slag. The prepared ore is mixed with concentrated sulfuric acid, and compressed air is blown through the reactor, as this accelerates etching. The reaction starts above 140 °C and is exothermic, which causes a further increase in temperature to 200–220 °C. The reaction mixture is left for 1–12 h to mature. The mixture is cooled to obtain a porous sinter, and then it is dissolved in water or diluted sulfuric acid. The process is performed at low temperature (below 85 °C), to prevent premature hydrolysis. The solution obtained, the so-called titanium liquor, is reduced with steel

scrap in order to transform Fe^{3+} into Fe^{2+} ions and partly also Ti^{4+} into Ti^{3+} ions. The solution is subjected to sedimentation, and the unreacted substrate is filtered out from the suspension. The content of TiO_2 in the product differs depending on the type of titanium ore: when ilmenite is used it is 8–12%, and when titanium slag is used it is 13–18%. When the temperature is decreased to 10 °C the main by-product of the process – hydrated iron sulfate ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$) – undergoes crystallization. When titanium slag with a TiO_2 content higher than 75% is used, the stage of iron sulfate crystallization is not necessary. The purified solution is concentrated to 200–250 g TiO_2/dm^3 . Hydrolysis of titanium sulfate is then performed (at 94–110 °C) by diluting the solution and maintaining the suspension at the boiling point. This process influences the size of the product particles and the degree of flocculation. The hydrated titanium dioxide formed is filtered off, washed with water, and subjected to whitening by hydrogen. The hydrated titanium dioxide obtained contains 20–28% sulfuric acid and sulfates. These impurities are removed by filtration

Inorganic, Hybrid, and Functional Pigments, Table 1 Characteristics of commonly known inorganic pigments

Pigment type	Production	Advantages	Disadvantages	Color
Carbon black	Decomposition of carbonaceous precursors	Rich color, color strength, light and weather resistant	Thickens paint, color limits application	Black
$\text{CH}_4 \xrightarrow{\text{temp.}} \text{C} + 2\text{H}_2$				
Titanium dioxide	Synthesized mainly by sulfate or chloride technology (alternatively using sol-gel, hydro- or solvothermal methods)	High opacity and color strength, relatively cheap, very good UV resistance, very high refractive index, good resistance to alkaline solutions, mostly chemically inert pigment	Tendency for agglomeration of primary particles, forms radicals that degrade the binder/polymer matrix	White
$\begin{aligned} \text{FeTiO}_3 + 2\text{H}_2\text{SO}_4 &\longrightarrow \text{TiOSO}_4 + 2\text{H}_2\text{O} + \text{FeSO}_4 \\ \text{TiOSO}_4 + 2\text{H}_2\text{O} &\longrightarrow \text{TiO(OH)}_2 + \text{H}_2\text{SO}_4 \\ \text{TiO(OH)}_2 &\xrightarrow{\text{temp.}} \text{TiO}_2 + \text{H}_2\text{O} \\ \text{TiO}_2 + \text{C} + 2\text{Cl}_2 &\longrightarrow \text{TiCl}_4 + \text{CO}_2 \\ \text{TiCl}_4 + \text{O}_2 &\longrightarrow \text{TiO}_2 + 2\text{Cl}_2 \end{aligned}$				
Iron oxides	Synthesized, obtained as minerals and also prepared from chemical wastes	Light and weather resistant, mostly unreactive	Cannot produce clean shades	Yellow, red, brown, black
$\begin{aligned} \text{C}_6\text{H}_5\text{NO}_2 + 2\text{Fe} + \text{H}_2\text{O} &\longrightarrow \text{C}_6\text{H}_5\text{NH}_2 + \text{Fe}_2\text{O}_3 \\ 3\text{Fe}_2\text{O}_3 + \text{H}_2 &\longrightarrow 2\text{Fe}_3\text{O}_4 + \text{H}_2\text{O} \\ 3\text{Fe}_2\text{O}_3 + \text{CO} &\longrightarrow 2\text{Fe}_3\text{O}_4 + \text{CO}_2 \\ 2\text{Fe}_3\text{O}_4 + 1/2\text{O}_2 &\xrightarrow{\text{temp.}} 3(\gamma\text{-Fe}_2\text{O}_3) \\ 2\text{Fe}_3\text{O}_4 + 1/2\text{O}_2 &\xrightarrow{\text{temp.}} 3(\alpha\text{-Fe}_2\text{O}_3) \end{aligned}$				
Zinc oxide	Chemical or metallurgical synthesis	Excellent heat stability, light and weather resistant, very good UV resistance, anticorrosive properties	Expensive chemical synthesis, tendency to form agglomerations	White
$\begin{aligned} \text{ZnSO}_4 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + \text{Na}_2\text{SO}_4 \\ \text{ZnCl}_2 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + 2\text{NaCl} \\ \text{Zn(CH}_3\text{COO)}_2 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + 2\text{CH}_3\text{COONa} \\ \dots\dots\dots &\dots\dots\dots \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{ZnSO}_4 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + \text{Na}_2\text{SO}_4 \\ \text{ZnCl}_2 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + 2\text{NaCl} \\ \text{Zn(CH}_3\text{COO)}_2 + 2\text{NaOH} &\longrightarrow \text{Zn(OH)}_2 + 2\text{CH}_3\text{COONa} \\ \dots\dots\dots &\dots\dots\dots \end{aligned}} \right\} \text{Zn(OH)}_2 \xrightarrow{\text{temp.}} \text{ZnO} + \text{H}_2\text{O}$				
$\text{ZnCO}_3 \xrightarrow{\text{temp.}} \text{ZnO} + \text{CO}_2$				

(continued)

Inorganic, Hybrid, and Functional Pigments, Table 1 (continued)

Pigment type	Production	Advantages	Disadvantages	Color
Zinc chromates	Synthesized	Anticorrosive properties	Thickens paints, dangerous synthesis with regard to reagent	Yellow
$4\text{ZnO} + 2\text{CrO}_3 + \text{K}_2\text{Cr}_2\text{O}_7 + \text{H}_2\text{O} \longrightarrow \text{K}_2\text{CrO}_4 + 3\text{ZnCrO}_4 \cdot \text{Zn}(\text{OH})_2$				
Ultramarine	Synthesized, seldom obtained as mineral	Rich color	Low stability in acidic environment	Blue
$\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4 + \text{Na}_2\text{CO}_3 + \text{Na}_2\text{SO}_4 \xrightarrow{\text{sulfur, carbon, temp.}} \text{Na}_{8-10}\text{Al}_6\text{Si}_6\text{O}_{24}\text{S}_{2-4} + n\text{CO}_2 + n\text{SO}_x + n\text{H}_2\text{O}$				
Chromium oxides	Synthesized as native pigments or produced from wastes	Light, weather, acid and alkali resistant, thermal stability	Does not give clear colors	Green
$4\text{CrO}_3 + 3\text{N}_2\text{H}_4 \longrightarrow 4\text{Cr}(\text{OH})_3 + 3\text{N}_2$ $2\text{Cr}(\text{OH})_3 \xrightarrow{\text{temp.}} \text{Cr}_2\text{O}_3 + 3\text{H}_2\text{O}$ $(\text{NH}_4)_2\text{Cr}_2\text{O}_7 \xrightarrow{\text{temp.}} \text{Cr}_2\text{O}_3 + \text{N}_2 + 4\text{H}_2\text{O}$ $\text{Na}_2\text{Cr}_2\text{O}_7 + 2\text{S} \xrightarrow{\text{temp.}} \text{Cr}_2\text{O}_3 + \text{Na}_2\text{SO}_4$				
Cadmium sulfides	Synthesized	High opacity and color strength	Poor weather resistance, very expensive	Greenish yellow, red, deep red
$\text{CdCl}_2 + (\text{NH}_2)_2\text{CS} + 2\text{NH}_4\text{OH} \longrightarrow \text{CdS} + (\text{NH}_2)_2\text{CO} + 2\text{NH}_4\text{Cl} + \text{H}_2\text{O}$				

and washing with water, followed by reduction in the process of whitening, in which most of impurities are removed. On completion of this process the content of colored impurities is of an order of ppm, while the content of sulfuric acid is still significant, at a level of 5–10%.

In the white part the hydrated titanium dioxide is again filtered, washed, and prepared for calcination. When the aim is to produce rutile, the nuclei of rutile and substances favoring transformation of anatase into rutile are introduced, and the process of calcination is performed at ca. 950 °C. When the aim is to obtain anatase, no admixtures are introduced and calcination is performed at a lower temperature. The hydrate formed in the process of calcination is subjected to calcination to remove water and to obtain TiO₂. When rotary filters are used the content of titanium dioxide in the product is 30–40%, but when pressure rotary filters or automated filtration press are used the content reaches 50%.

A schematic presentation of the sulfate method of titanium white production is given in Fig. 4, and the reactions at each step are listed in Table 1.

Chloride Process

In the chloride process the substrates are rutile or enriched titanium ore containing over 90% TiO₂. This method is more modern, more technologically advanced and less burdensome to the natural environment; however it enables the production of rutile only. The problems related to this method are the high temperature of the process; the aggressive medium; the toxicity of chloride, carbon monochloride, and titanium tetrachloride; and the risk of uncontrolled emission of chlorine gas. The advantages of this process are that the pigments obtained have higher resistance to light and are brighter than those obtained by the sulfate method.

The substrate introduced into the reactor is maintained in the fluid phase by a flux of air. The substrate is heated to 650 °C and then refined coke is introduced in the amount of 250–300 kg per tonne of TiO₂. The substrate should be dry and should not contain too small particles that would be carried by the stream of forming TiCl₄ and could contaminate the product. Burning coke causes an increase in temperature to about 1000 °C. In these conditions the introduction of

chlorine and continuous supply of titanium ore (substrate) and coke begin.

The reaction leads to the formation of titanium tetrachloride gas, which is the main product, and by-products including chlorides of contaminating metals, e.g., iron, manganese, and chromium, which together with the unreacted substrate are separated from the reaction mixture by fractional condensation. The formed titanium tetrachloride and the accompanying chlorides are distilled off from the fluid bed. These gases are then cooled with liquid TiCl_4 . Some of the metal chlorides (mainly iron chloride) can be deposited on the surface of TiCl_4 upon crystallization. To prevent this, the chlorides are condensed off. In this way the impurities of TiCl_4 are removed. Cooling is carried out in stages. First the gases are cooled to a temperature below 300°C . Then the chlorides co-occurring with titanium tetrachloride are separated by condensation or sublimation. After this stage the gas is mostly composed of TiCl_4 and is cooled to a temperature below 0°C , at which condensation of TiCl_4 takes place. The raw titanium tetrachloride is carefully purified, at first by the chemical method to separate the substances that could not be removed upon distillation, e.g., VCl_4 , VOCl_3 , and then by vacuum distillation, which gives highly pure TiCl_4 . The pure titanium tetrachloride is evaporated and combusted in a stream of oxygen or air to obtain TiO_2 and chlorine. The reaction is weakly exothermic and needs a high temperature in the range $900\text{--}1400^\circ\text{C}$. The pigment obtained with a mixture of gases (Cl_2 , O_2 , CO_2) is cooled and then separated by an indirect or direct method. The stream of gases is recirculated to the cooling zone of the furnace and to the chlorination stage.

A schematic presentation of the chloride process of titanium white production is given in Fig. 4, and the reactions at each stage are listed in Table 1.

After cooling, titanium dioxide obtained via either the sulfate or chloride method is ground using a wet or dry method and dispersed in water. It can be subjected to surface modification with hydrated oxides, e.g., silica, alumina, or zirconia, to obtain varieties of pigment for special applications. The main aim of the inorganic treatment is to improve the pigment's stability,

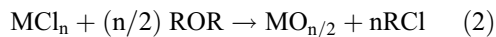
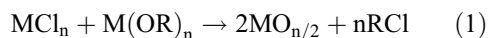
reactivity, and its dispersibility. The treatment is carried out in the water phase, and solutions of substances used to modify the particles of titanium dioxide are added in a defined sequence or simultaneously. By regulating the pH of the suspension, hydrated oxides are precipitated on the surface of the pigment particles. The form of the oxides strongly depends on the conditions of precipitation. Alumina, silica, zirconia, or other oxides do not occur in the pigment as an admixture, but as a molecular layer bonded to the pigment surface. After the treatment the pigment is filtered again, dried, and grounded. The product is packed and stored.

Other alternative methods of TiO_2 production, which allow the control of process parameters as well as the physicochemical and useful properties of the products, include the sol-gel, hydro- and solvothermal methods, and chemical vapor deposition.

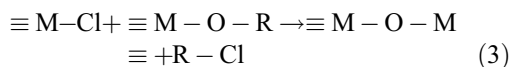
Sol-Gel Method

In a typical sol-gel process, a colloidal suspension or sol is formed as a result of hydrolysis and polymerization of precursors, which are usually inorganic metal salts or organometallic compounds such as metal alkoxides. Polymerization and loss of solvent lead to conversion of the liquid sol to a solid state gel phase.

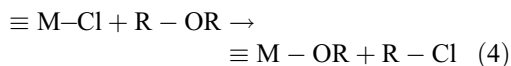
The sol-gel method is based on hydrolysis and condensation of metal alkoxide or metal salt. The main process is the reaction of the metal chloride with metal alkoxide or organic ether acting as oxygen donors, according to the Eqs. (1, 2):



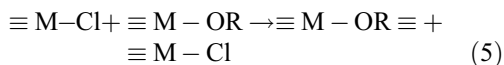
In these reactions, M–O–M bond formation is promoted by condensation between M–Cl and M–OR, according to Eq. (3):



In the reaction with ether (4) metal alkoxide is formed as a result of alcoholysis with M–Cl:



At room temperature these reactions are slow, and oxide formation is usually favored by an elevated temperature in the range 80–150 °C. The main reaction (5), between metal chloride and metal alkoxide, takes place at room temperature.



An important advantage of the sol-gel method is the possibility of obtaining TiO₂ particles of nanometric size and controlled shape.

Hydrothermal and Solvothermal Methods

The method of hydrothermal synthesis is commonly used for obtaining titanium dioxide with small-diameter particles for the ceramic industry. TiO₂ nanoparticles are obtained during hydrothermal processing by peptization of titanium precursors with water. The hydrothermal method is commonly used for the production of different morphological forms of TiO₂ such as nanotubes, nanowires, and nanorods.

The solvothermal method is very similar to the hydrothermal method – the difference is that in the former an anhydrous solvent is used instead of water. The range of temperatures can be much greater than in the hydrothermal method, and depends on the boiling point of the organic solvent used. In the method of solvothermal synthesis it is easier to control the size, shape, and crystallinity of TiO₂ nanoparticles than in the hydrothermal case. The solvothermal method is regarded as a versatile method for the synthesis of nanoparticles with narrow particle size distribution.

Chemical Vapor Deposition

Chemical vapor deposition (CVD) is a method for the thermo-chemical processing of materials. The idea of the method is to deposit thin films on a given material to change its physical, chemical, or mechanical properties. The reaction chamber is

charged with titanium precursors, mostly in the gas phase; there on a hot substrate the chemical reactions take place. Usually CVD methods require high temperatures of 900–1100 °C or higher, necessary for film formation, and this condition significantly restricts the use of these methods. In order to obtain the target products, different gas or liquid substrates (precursors) are used, e.g., halogens, chlorides, carbonyls, and volatile organic compounds of titanium.

The wide range of production methods, as well as the outstanding physicochemical and utilizable properties of TiO₂, makes this pigment an extremely attractive product in the paint and lacquer industries.

The greatest amount of titanium white is used for the production of paints. Thanks to a very high light refraction index the paint coatings containing titanium dioxide are intensely white and have high opacity, which means that only small amounts of TiO₂ are needed to produce an excellent white opaque coating. Titanium white also increases the coating durability and ensures high covering power; this means that the coating can be thinner, which improves its functional properties. This pigment is used to obtain high quality lacquers, both white and in pastel colors in mixtures with other pigments and emulsion paints. Paint production uses a large variety of titanium white pigments: rutile, anatase, pigments with modified and unmodified surface, or pigments modified for special needs. The type of pigment depends on the type of paint and its application. For external use (paints exposed to sunlight) rutile varieties are used, while for indoor applications anatase is preferred. Anatase varieties with lower resistance are used in cheap dispersive paints, in self-cleaning paints and in paints for road markings. Rutile varieties have much wider use and can be applied in all types of paints.

TiO₂ is also used for the production of printing inks; less than 3% of global production of titanium white is used for this purpose. TiO₂ pigments in printing inks ensure opacity and high brightness of the coating. Because of the thinness of printing coatings, TiO₂ pigments used for this purpose should have a narrow particle size distribution and should not contain agglomerates. The proper

choice of pigment is very important as it affects many properties of the ink, such as luster, color, covering power, susceptibility to sedimentation, wearing, and rheological properties. The varieties of titanium white pigment should be chosen depending on the final intended use of the ink.

The second best known and widely used white pigment is *zinc oxide* (C.I. Pigment White 4; ZnO). This pigment is produced especially by a metallurgical process based on zinc ore calcinations [8].

Metallurgical Process

Zinc oxide products are divided into:

- Type A – obtained by a direct process (American process)
- Type B – obtained by an indirect process (French process)

In the direct (American) process, zinc ore is reduced by heating with the addition of coal (e.g., anthracite) and then oxidation of zinc vapors in the same reactor in one production cycle. This process was proposed by Samuel Wetherill.

In the indirect (French) process, metallic zinc is molten and evaporated at about 907 °C. Immediate reaction of zinc vapors with oxygen from the air gives ZnO. Zinc oxide particles are transported through a cooling channel and collected at a station of cyclones and sock filters. This process was popularized by LeClaire (France) in 1844, and since that time it has been called the French process. The product is obtained in the form of agglomerates of mean size from 0.1 to a few μm . The particles of ZnO have mainly spheroidal shape. The French process is carried out in vertical furnaces. Zinc oxide type B shows higher purity than type A.

Of great importance in the technology of production of ZnO with defined physicochemical and structural properties are the chemical processes described below.

Controlled Precipitation

Controlled precipitation is a widely used method for the synthesis of zinc oxide, as it enables the

obtaining of products with reproducible properties. The idea of the method is to perform a fast and spontaneous reduction of zinc salt with a reducing agent that would stop the growth of particles at a certain size, followed by precipitation of ZnO precursor from the solution. At the next step the precursor is subjected to thermal treatment and grinding. Breaking up of the agglomerates formed is very difficult, so the calcined powders show a high level of agglomeration. The process of precipitation is controlled by such parameters as pH, temperature, and time of precipitation. The most important zinc precursors in chemical processes are zinc acetate and zinc hydroxide. A necessary stage of this process is the calcination of hydrated ZnO at above 120 °C (see Table 1).

Mechanochemical Process

The mechanochemical process is a cheap and simple method for the synthesis of nanoparticles on a large scale. It is based on high-energy dry milling, which initiates a reaction upon ball-powder collisions in a ball-mill at a low temperature. The main difficulty in this method is the uniform grinding of the powder and reduction of grains to a target size. This size decreases with increasing time and energy of milling. Unfortunately, a longer time of milling leads to the appearance of increasing quantities of impurities. The benefits of this method are the low cost, small particles, and limited tendency to form agglomerations, as well as the highly uniform crystalline structure and morphology.

The initial materials used in this method are mainly anhydrous ZnCl_2 and Na_2CO_3 . The reactor is charged with NaCl, which is the reaction medium and which separates the forming nanoparticles. The zinc precursor formed – ZnCO_3 – is subjected to calcination at a temperature in the range 400–800 °C.

Other significant methods for the synthesis of zinc oxide are the hydro- and solvothermal methods and synthesis of morphologically defined particles in emulsion and microemulsion systems, the same as in the case of TiO_2 [8].

Besides TiO_2 and ZnO, commonly known white pigments are *lead white* (C.I. Pigment

White 1; $2\text{PbCO}_3\text{Pb(OH)}_2$) and *zinc sulfide* (C.I. Pigment White 7; ZnS).

Black pigments – the optical effect of these is achieved by nonselective light absorption; among this group of pigments the most common are carbon black and black iron oxide.

The most often used black pigment is *carbon black* (C.I. Pigment Black 6, 7, 8). The majority of carbon black types used as pigments have highly developed surface area, which means that their oil absorption numbers are high. Their high surface activity is responsible for poor wettability, the effect of flotation, and adsorption of paint components on the surface. Carbon black is an inorganic pigment, but it has many properties of organic ones. It is composed of carbon. The wide range of pigments depends on the types of their precursors and the methods of their production. The paint industry uses only modest amounts of black pigments when compared to the rubber industry. In the production of rubber for tires, carbon black not only provides color, but also increases mechanical strength. Black pigments based on carbon black have excellent resistance to light and solvents, although certain solvents can extract some colored contaminants from cheap types of pigments. Black pigments show excellent chemical and thermal stability. The intensity of black depends on the size of the particles: the smaller they are, the more intense the color.

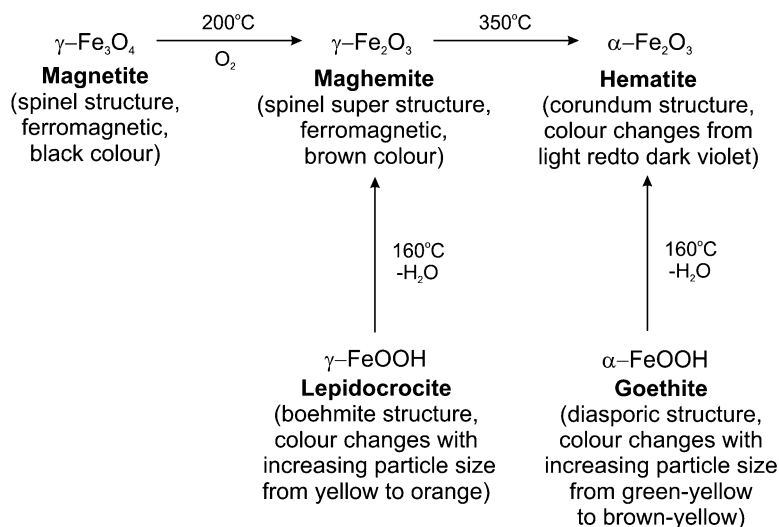
A very attractive black pigment is also *iron oxide* (C.I. Pigment Black 11; Fe(II) and Fe(III) oxide; FeO , Fe_2O_3). In natural form it is found as magnetite (also maghemite, hematite, goethite, lepidocrocite – Fig. 5), which can be used as a pigment, although most iron oxide black pigments are synthesized. Iron oxide is a cheap inert pigment of excellent chemical resistance, insoluble in organic solvents and with excellent durability, though unfortunately it also has limited thermal stability and low tinctorial strength. This group of pigments also includes *graphite* (C.I. Pigment Black 10), formed by crystallization of carbon in hexagonal form.

Color pigments – the optical effect of these is achieved by selective light absorption often in combination with selective scattering; examples include brown, yellow, red, blue, and green pigments.

Brown pigments, of which an example is brown iron oxide (C.I. Pigment Brown 6 synthetic and 7 natural; Fe_2O_3), naturally occurs as burnt sienna or burnt umber. Different shades can be obtained depending on the presence of impurities, in particular manganese oxide. These pigments are more often used for artists' paints than for commercial paints. They have low tinctorial strength. The synthetic pigments give an intense brown color with excellent resistance. They are not widely used for paint production, as similar

Inorganic, Hybrid, and Functional Pigments,

Fig. 5 Typical forms of iron oxide-based pigments



shades can be obtained by mixing cheaper pigments.

Lead chromates (C.I. Pigment Yellow 34, Orange 21; $\text{PbCrO}_4/\text{PbSO}_4$; orange $\text{PbCrO}_4 \cdot \text{Pb}(\text{OH})_2$) are a good example of yellow pigments. The pigments can be composed of pure lead chromate or of a mixture of lead chromate and lead sulfate. They give colors from green-yellow to orange. Chrome yellow displays excellent opacity, a low oil absorption number, high brightness of shades, and high color saturation, which make it an ideal pigment for the production of paints in a wide gamut of yellow shades. Yellow pigments have very high resistance to solvents and relatively good thermal stability, which can be improved by chemical stabilization. They are sensitive to alkaline compounds and to acids, which cause them to fade. Their resistance to light is usually satisfactory, although they can become slightly darker under exposure to light. These drawbacks can be counteracted by surface modification with such materials as silica, antimony, aluminum oxide, or different metals. Other yellow pigments are cadmium yellow (C.I. Pigment Yellow 37, 35 lithopone; CdS ; lithopone – CdS containing ZnS) and yellow iron oxide (C.I. Pigment Yellow 42 synthetic, 43 natural; $\text{Fe}_2\text{O}_3 \cdot \text{H}_2\text{O}$, or more exactly $\text{FeO}(\text{OH})$).

Excellent functional properties, such as thermal and chemical stability, and resistance to solvents are provided by red iron oxides (C.I. Pigment Red 101 synthetic, 102 natural; Fe_2O_3), which belong to the category of red pigments. They are also highly opaque and have quite

low tinctorial strength. Despite the serious drawback of its dull brown-red shade, it has the great advantage of low cost.

Ultramarine (C.I. Pigment Blue 29, Violet 15; general formula $\text{Na}_7\text{A}_{16}\text{Si}_6\text{O}_{24}\text{S}_3$), which belongs to the group of blue pigments, has attractive features, including dazzling color and pure shade that can vary from pink, through violet to green. It is also characterized by excellent thermal stability, resistance to solvents, and good resistance to light and alkaline compounds. It is not resistant to acids, and thus it cannot be used in exterior paints. Other well-known blue pigments are Prussian blue (C.I. Pigment Blue 27; $\text{FeK}_4\text{Fe}(\text{CN})_6$) and cobalt blue (C.I. Pigment Blue 36; $\text{Co}(\text{Al,Cr})_2\text{O}_4$).

Last in sequence but not least in importance are the green pigments, such as green chromium oxide (C.I. Pigment Green 17; Cr_2O_3) and hydrated chromium oxide (C.I. Pigment Green 18; $\text{Cr}_2\text{O}(\text{OH})_4$). These pigments have an opaque shade, good opacity but low tinctorial strength. They also offer excellent thermal stability, and chemical resistance to solvents and light [2, 9, 10]

Because of their specific properties, inorganic pigments are used mainly in the paint and lacquer industries. However, in order to meet set requirements for the final products (paints and coatings), they must be formulated with other additives which do not provide any staining or opacity. These additives are known as extenders (Table 2). They may be used to change typical pigment properties as well as specific functions [3, 11].

A special group of inorganic pigments consists of *pigments with luster (mirror effect)* and

Inorganic, Hybrid, and Functional Pigments, Table 2 Example of pigment extenders

Extender type	Formula	Uses
Calcium carbonate	CaCO_3	Additive for paints containing solvent, printing inks
Silica	SiO_2	Additive for rheology and thixotropy control, anti-settling, thickening and anti-sagging agent
Talc	$\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$	Gives high hiding power and a stain finish, provides a matting effect and decreased moisture permeability
Barytes	BaSO_4	Additive for waterborne paint systems
Kaolin	$\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$	Decreases viscosity
Mica	$\text{AB}_{2-3}(\text{Al,Si})\text{Si}_3\text{O}_{10}(\text{OH})_2$; A = K, Na, Ca; B = Al, Fe, Mg, Li	UV and chemically resistant, improves water resistance, provides pearl effect

luminescent pigments as well as **hybrid pigments** [12–16]. The first group of these special pigments includes the following:

- **Metallic pigments** (only inorganic) contain lamellar metal particles of high refraction index, e.g., flakes of copper, gold, aluminum.
- **Pearlescent pigments** contain transparent lamellar particles, responsible for a pearl shine, which is the result of multiple reflections from particles oriented in parallel; examples include specially treated alkaline lead carbonates, bismuth oxychloride, and titanium dioxide supported on mica.
- **Interference pigments** – here the optical effect is a result of the thickness of the oxide layer deposited on supports (similarly as in pearlescent pigments); examples are oxides of iron or chromium supported on mica; pigments in this group also display opalescence.

Also very important are luminescent pigments, including:

- **Fluorescence pigments**, of which the optical effect is based on selective absorption of light and simultaneous luminescence, and is initiated by high-energy radiation from the ultraviolet or shortwave visible range; examples are radioactive luminescent pigments.
- **Phosphorescence pigments** (only inorganic) – here the optical effect is based on selective light absorption and scattering followed by delayed luminescence, and is also initiated by high-energy radiation from the ultraviolet or shortwave visible range; examples are zinc sulfide and alkaline earth metal sulfides “dotted” with heavy metal ions.

Hybrid Pigments

Recently, much interest has been paid to the synthesis of hybrid pigments. They can occur in four main combinations: organic/organic, inorganic/inorganic, organic/inorganic, and organic/bioorganic. This group of pigments includes those obtained on the basis of organic dyes and oxide supports including among others silica, titania, or alumina. Substances of this type are

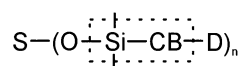
obtained mainly as a result of adsorption of organic molecules on the synthetic or mineral oxide support. Adsorption of organic compounds such as organic dyes is determined first of all by the type of interactions (covalent bonds, electrostatic force, hydrogen bonds, etc.) and the character of the inorganic support (SiO_2 , TiO_2 , $\text{SiO}_2\text{-TiO}_2$, Al_2O_3 , aluminosilicates and carbonaceous materials), among other things its surface properties and dispersion abilities [14–27].

The most often used selective adsorbent of organic dyes is silica synthesized by the sol-gel method. Alternative supports for SiO_2 obtained by this method are the silicas synthesized in the precipitation reaction from water solutions of lithium group alkali metal silicates, mainly sodium silicate. A large group of selective adsorbents or supports consists of mineral substances such as kaolin and montmorillonite. The core of pigment composites can also consist of biopolymers of natural origin (chitin, chitosan, lignin) or synthetic ones.

Hybrid pigments of this type can be obtained in different colors depending on the color of the initial dye. They show many advantages over conventional pigments. The particle size of such pigments depends on those of the initial support. Moreover, from the same core it is possible to get pigments with the same particle size, density and surface properties, but differing in color and hue.

Pigments Obtained by Grafting of an Organic Dye to a Modified Oxides Surface

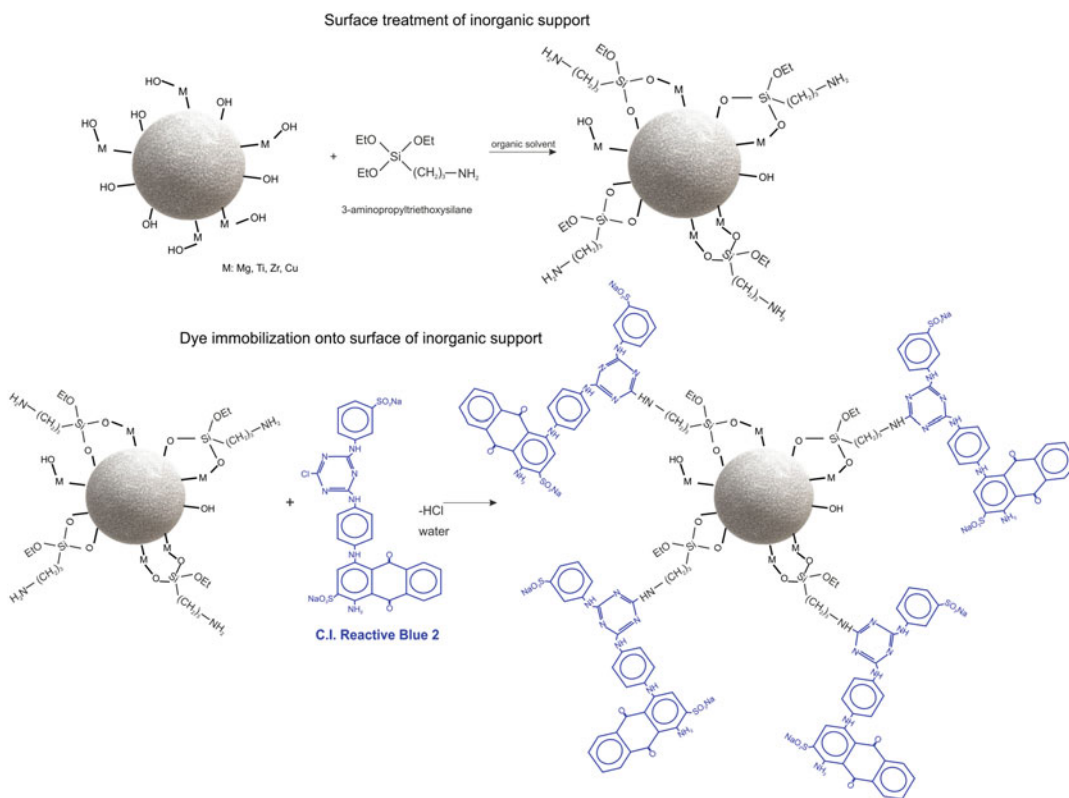
Such pigments are composed of a core that is a fine powder – SiO_2 , TiO_2 , Al_2O_3 , etc. – and an organic dye chemically bonded with the oxide surface by a coupling agent, e.g., aminosilane. They show high color strength, excellent durability, and resistance to solvents, water, temperature, and light. Their structure can be described by the formula:



where:

S is the support;

CB is a coupling bridge – an alkyl group (3–6 carbon atoms) containing –NH– or –NHCO– in a chain like (for example):



Inorganic, Hybrid, and Functional Pigments, Fig. 6 Mechanism of hybrid pigment formation by adsorption of organic dye onto aminosilane-functionalized

oxide support. (Created based on [28] with permission, Copyright 2013: Elsevier Publisher)

$-\text{CH}_2\text{CH}_2\text{CH}_2-$, $-\text{NH}-\text{CH}_2\text{CH}_2-\text{NH}_2$,
 $-\text{CH}_2\text{CH}_2\text{CH}_2-\text{NHCO}-$, $\text{R}-\text{N}=\text{CH}-$,
 $\text{R}-\text{NH}-\text{CH}_2-$ or $\text{R}-\text{NHSO}_2-$,
 D is the organic dye.

The most popular inorganic support for hybrid pigment formulation is silica. Hydrophilic silicas are usually colorless and have silanol groups on the surface. These species can react with many functional groups through covalent bonds. The process of dyeing of these silicas is multistage. At first silica reacts with a coupling agent in order to bind a compound selectively reacting with a given dye, and then the dye (reactive, acidic, or basic) reacts with the coupling agent (e.g., silane) to give silica particles with covalently bonded dye. The dye permanently bonded to the silanol coupling agent cannot be leached by a solvent, which leads to reduced toxicity of the components of a paint or ink.

The sequence of reactions in this method is shown schematically in Fig. 6. The sequence includes modification of the oxide surface with 3-aminopropyltriethoxysilane in order to introduce amino groups, which in the subsequent process react with the reactive dye (e.g., C.I. Reactive Blue 2) to give dyed silica particles [28].

Through simple and economical processes it is possible to obtain colored particles used as pigments for inks. These inks show high resistance to water and temperature, and have lower toxicity. The inks can be made in a wide gamut of colors and can be used for printing on foil, smooth or coated paper.

This process can be also realized by multistep synthesis of a chromophore on the surface of an inorganic support.

The pigment synthesis involves modification of the support surface – which becomes the core

of the pigment – with an aminosilane coupling agent to introduce surface amino groups and then bonding the dye with amino groups. For this purpose silicas of small diameters are used, as when silica particles are small the area of contact is usually larger, which is beneficial for mixing with other materials.

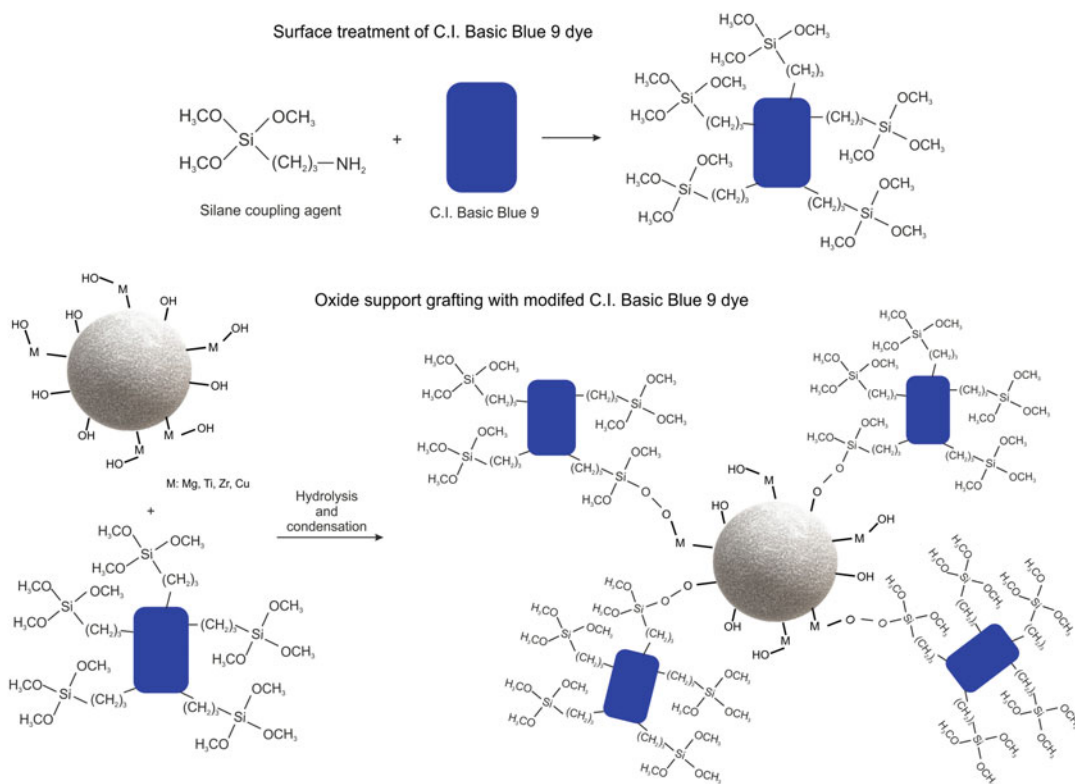
Pigments Obtained in the Process of Silica Synthesis

Pigments of this type are obtained with silica (or other oxide material) synthesized in a process of hydrolysis and condensation of tetraethoxysilane (TEOS) in a mixture of ethanol/water/ammonia and in the presence of cationic dyes such as C.I. Basic Blue 9 (Fig. 7), additionally treated with, e.g., silane coupling agents in order to increase its adhesion to oxide material surface.

The initial dye solution is prepared in a water medium and is then filtered off through a suitable membrane. In order to synthesize inorganic particles containing a dye, the silica sol and dye solution are mixed to obtain a permanent product. Often the whole amount of the dye used in the process is incorporated into the structure of the pigment, as indicated by a colorless solution remaining after the process. The color particles are filtered off and washed with distilled water until no dye is left in the filtrate.

Retention of dye in the support particles is tested by dispersion in ethanol or acetone, followed by ultrasound treatment and centrifugation. This procedure is repeated a few times to check the stability of the bonding of the dye to the support surface.

There is also interest in combining the functional features of inorganic pigments (TiO_2) or their derivatives ($\text{TiO}_2\text{-SiO}_2$) with selected dyes



Inorganic, Hybrid, and Functional Pigments, Fig. 7 Modification of the pigment and binding with the silica surface

used in the food and/or pharmaceutical industries (e.g., C.I. Food Yellow 4, C.I. Food Blue 5:2, C.I. Food Red 9) [16]. An example mechanism of synthesis of such a hybrid pigment is illustrated in Fig. 8.

Such a combination makes it possible to obtain a homogeneous hybrid system, which is a great advantage from the application point of view, as separate introduction of dye and pigment at the stage of drug formulation (e.g., tablets) is no longer needed and problems with their steric and electrokinetic stability are eliminated.

Colored Bioorganic-Based Hybrid Pigments

Besides various type of inorganic materials used as supports of organic dyes there is also a valuable group of naturally occurring biomaterials such as spongin-based systems. Among them, there are different kinds of organic substances, which skeleton of some sponges can be built of: chitin, spongin, and collagen. Spongin is a biopolymer of still unknown chemical structure, but seems to be a naturally occurring hybrid between collagen and keratin-like proteins. From a practical point of view, the properties and architecture of spongin play a key role and are consequence of its structure. The organic chemistry of sponges is also the focus of intense research because they have been known to contain active biomolecules, which are of therapeutic and pharmaceutical value. Thousands of sponge-derived bioactive metabolites have been isolated and identified so far [29, 30]. Most bioactive compounds from sponges belong to anti-inflammatory, antitumor, immuno- or neurosuppressive, antiviral, antimalarial, antibiotic, cytotoxic, cytoprotective, enzyme-inhibitory, or antifouling class. That is why, the utilization of materials of natural origin like structural proteins – spongin, has been gaining increasing scientific attention. Key features contributing to the popularity of these biomaterials include biodegradability, ecological safety, low cost, high compatibility with the environment and renewability.

Recently, natural dyes are widely studied compounds in modern science and new information about their anticancer, anti-inflammatory, and antioxidant properties are published

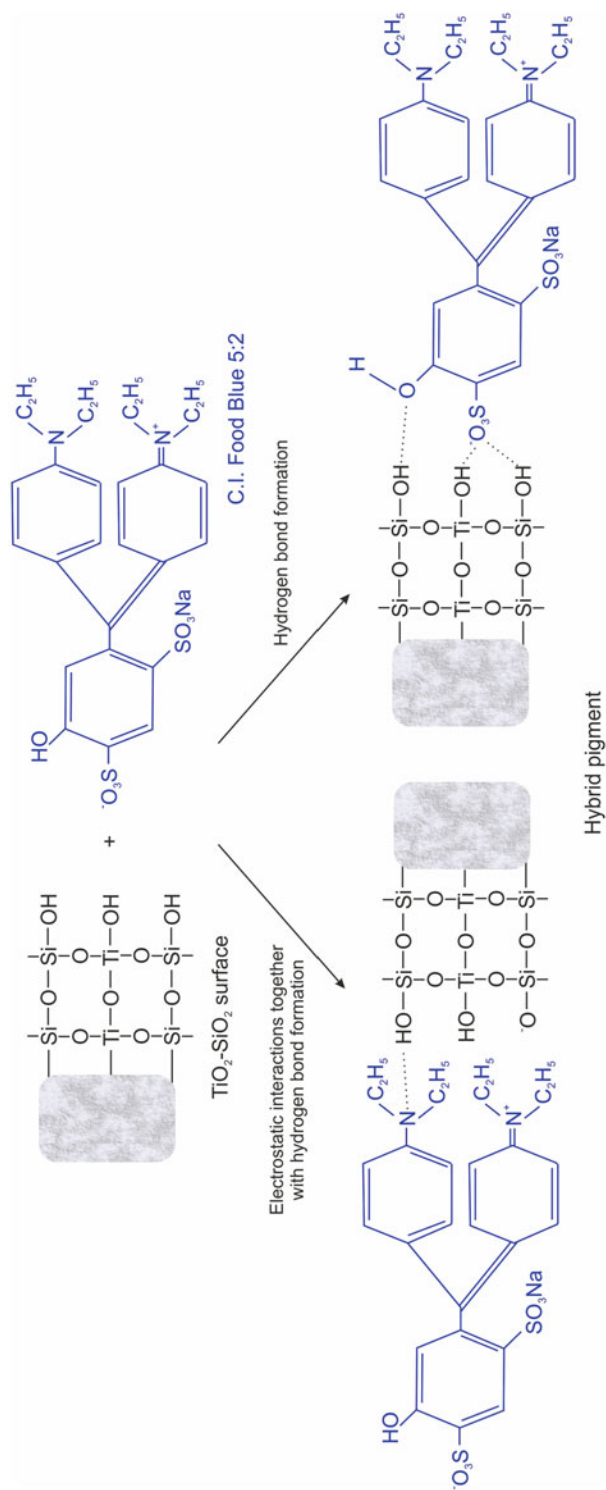
[31, 32]. Natural colorants are usually obtained by means of aqueous extraction of plant material; the use of solvents is less common as substantial amounts of plant material have to be extracted. As a result of the nonselective extraction process, natural colorants usually contain a number of different color-providing chemical substances. The classification of a dye plant or animal source to a certain group of colorant usually is based on the properties of the most important species of dye present in the extract.

Natural colorants have some disadvantages as compared to synthetic ones, as they are more expensive and less stable, susceptible to degradation affected by light, pH, temperature, sulfite, ascorbic acid, and enzymes [33]. Natural dyes are renewable and sustainable bioresource products with minimum environmental impact, biodegradable, and nontoxic [34]. They find applications in food and textile coloration, as pH indicators, in cosmetics, pharmaceuticals, or in dye-sensitized solar cell production [35]. However, the most important limitation reported is the low bio-distribution and bioavailability as well as instability. Therefore, adsorption of these natural compounds onto biopolymers support improves their stability and enables use of their antibacterial and antiradical properties in further applications.

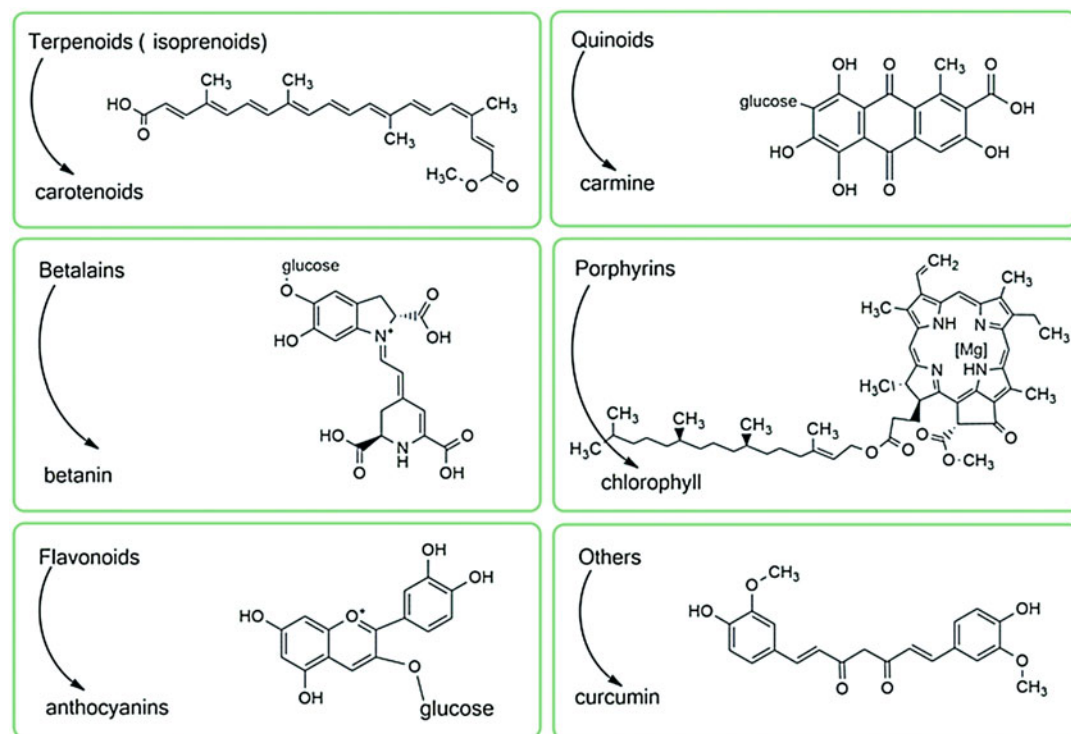
There are different groups of natural dyes: anthraquinones, flavonoids (including anthocyanins), or porphyrins. According to their chemical structure they can be divided into several classes, presented in Fig. 9.

Carmine, also called C.I. Natural Red 4 or cochineal, is a dark red dye of animal origin [36]. It can prevent coronary artery disease, plays a role in Alzheimer's disease treating, exhibits cancer chemopreventive activity, and is used in drug delivery products [37].

Anthocyanins are a group of plant dyes belonging to the flavonoids, with a polyphenol structure. Different factors affect the color and stability of these compounds, thus their molecules are highly unstable and easily susceptible to degradation under the influence of pH, temperature, ionizing, UV and solar radiation, the presence of copigments, metallic ions, enzymes, oxygen, and



Inorganic, Hybrid, and Functional Pigments, Fig. 8 Mechanism of C.I. Food Blue 5:2 interactions with $\text{TiO}_2\text{-SiO}_2$ surface (Created based on [16] with permission, Copyright 2013; Elsevier Publisher)



Inorganic, Hybrid, and Functional Pigments, Fig. 9 Types of natural dyes with the examples

ascorbic acid [33]. Anthocyanins have electron-donating abilities hence they have great potential for application as sensitizers in photoelectron-chemical dye-sensitized solar cells [38]. In food, cosmetics, and medicines they serve not only as a colorant but also their wide range of bioactive properties and health benefits is exploited. The antioxidant, antibacterial, anti-inflammatory, antiviral, antifungal, chemopreventive, anticancer, and antimutagenic properties of anthocyanins are well known and described in literature [39, 40].

Chlorophyllin, a porphyrin-like structure dye of plant origin is a precursor of chlorophyllin whose stable macrocycle structure affect the relative thermal stability. This above-mentioned feature as well as water solubility (in contrast to chlorophyll), nontoxicity, and numerous bioactive properties cause the great potential of application for chlorophyllin. Recent studies reveal antioxidant, antimutagenic, and anticarcinogenic properties, because of the ability to bind a planar

compounds such as heterocyclic amines, dioxin, aflatoxin, and benzo[a]pyrene [41, 42]. It also should be mentioned that sodium copper chlorophyllin is used as a dietary supplement, in food, drugs and cosmetics, in textile dyes, as an internal deodorizer and as a natural wound healer [43, 44].

Other porphyrin-like structure dyes include phthalocyanines. Because of their structure they exhibit unique properties: high redox activity; high molar absorption coefficient (light-absorbing electron donors); thermal stability; oxidant, acid, and alkali resistance as well as high reactivity. These features enable to use phthalocyanines in electronics and sensor application, as a dyes and catalysts or for solar cell production [45].

Special type of phthalocyanines include metal phthalocyanines, which exhibit high catalytic activity (in a so-called biomimetic catalytic system) even at ambient conditions, moreover they are relatively easy to synthesize and commercially

available [46, 47]. However, one limitation of these compounds is that they are inconvenient to use, since they are generally available only as powder, in solution (which hinders their separation from the solution and catalyst recycling) or in the form of thin film [48, 49]. A better strategy for their practical use might be to attach them to a suitable support material. Adsorption of metal phthalocyanines on a solid support is an effective way to remedy the drawbacks of the homogeneous catalyst and enable the creation of a heterogeneous system (possibility of reutilization, increase in the surface area of the catalyst).

Mentioned dyes can be easily adsorbed onto the surface of different supports including functional biopolymers. In comparison with other methods this process is fast, facile, takes place in mild conditions, and is broadly applied for dyes. That is why, the preparation of the spongin/dye hybrids allows to combine the functional properties of the dyes with the thermally and mechanically resistant natural support, and enables designing of products with unique physicochemical properties and the possibility of finding interesting application, especially when using natural and synthetic dyes with such properties as: antioxidant, antibacterial, and catalytic.

Preparation of such spongin/dyes systems require specific process parameters control including concentration of the adsorbate solution, the pH and the contact time of the reagents. The higher dye concentration and duration of adsorption, the higher amount of dye adsorbed onto spongin should be expected. Nevertheless, the crucial parameter is pH. This is related to the presence of different forms of dyes depending on the different pH values (see Fig. 10). Moreover, pH affects the nature of adsorbent/adsorbate interactions.

The presence of additional ions (Na^+ and Cl^-) is also a key aspect that affects the efficiency of dyes adsorption onto spongin supports. According to the surface chemistry theory, if there is an attractive electrostatic interaction between adsorbate and adsorbent, then an increase in ionic strength should cause a decrease in the

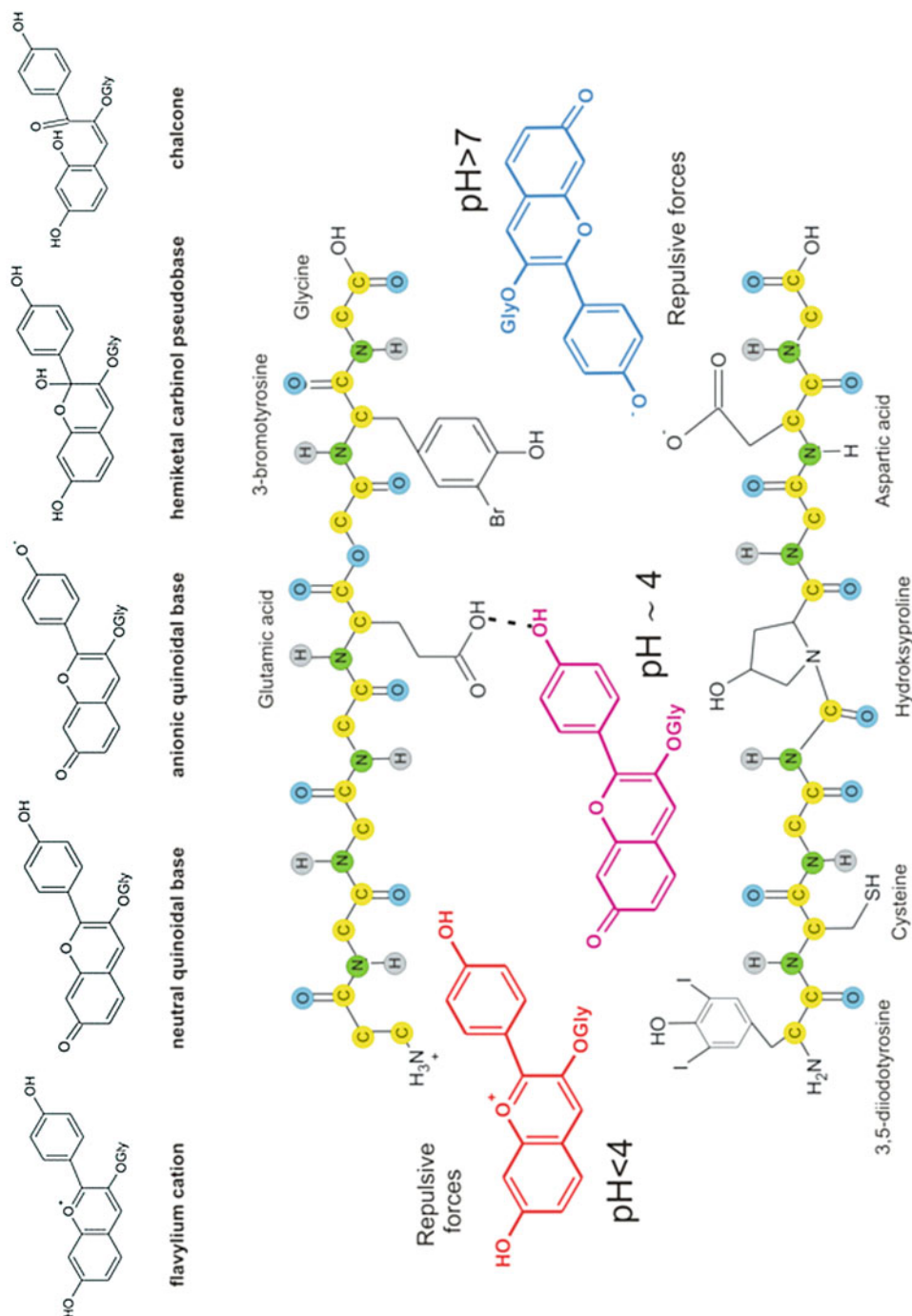
adsorption capacity of the adsorbent. The converse is found in case of repulsive interaction; increasing ionic strength affect an increase in adsorption capacity, especially when the concentration is high. The presence of additional ions from salt in the medium decreases the repulsion between adjacent dye particles, allowing the adsorbed molecules on the surface to be closer to each other [51, 52].

The realization of the adsorption process, e.g., of copper phthalocyanine onto spongin is promoted by acidic pH, which affect the interaction between dye and spongin skeleton. CuPC have in their structure functional groups (sulfonic) capable for dissociation. Spongin, as a protein material is built of several amino acids with characteristic amino groups ($-\text{NH}_2$), which could be protonated into cationic form ($-\text{NH}_3^+$). Therefore, the interaction between oppositely charged functional groups is possible (Fig. 11).

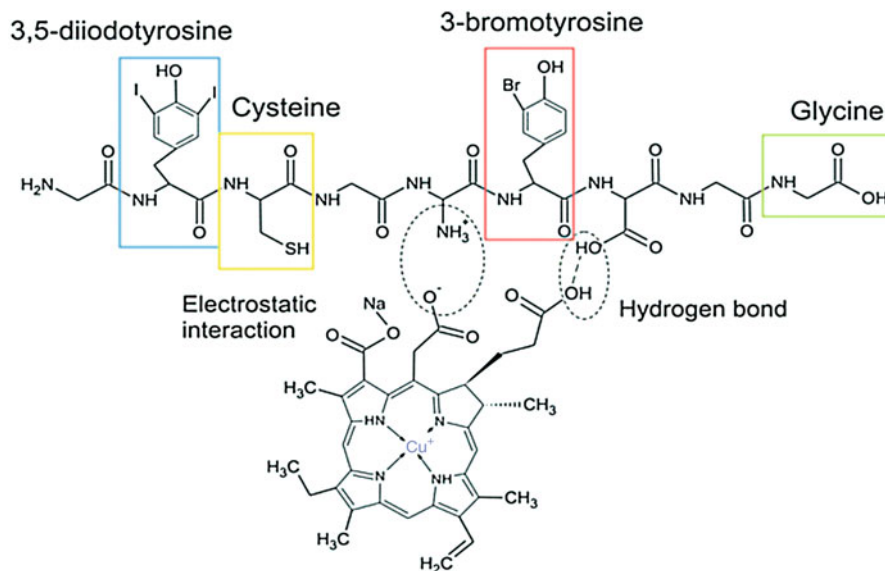
The differences in the amount of adsorbed dyes may be caused by many factors. Attention should be given to the chemical structure of the dye, its molecular weight or the number, and reactivity of functional groups.

Because of its nontoxicity and antiradical properties naturally occurring anthocyanin has a potential for the utilization as a harmless coloring material. Its color instability under, for example, light irradiation limits its practical use. To the wider use of the anthocyanin, the immobilization on the inorganic host for an easy handling as well as the improvement of the stability is required. The impregnation of spongin skeleton with antioxidant substances could have a positive effect in both application areas, due to free radical scavenging properties, as in the case of anthocyanin dye. The adsorbed anthocyanin exhibits antioxidant properties at a level compared with dye solution (Fig. 12). This could provide a better chemical system in terms of stability without loss of bioavailability.

Numbers of bioactivities are also ascribed to chlorophyllin such as, e.g., antibacterial properties. The basic advantage of the spongin/chlorophyllin systems, in comparison to pure dye should be its insolubility, which provides the

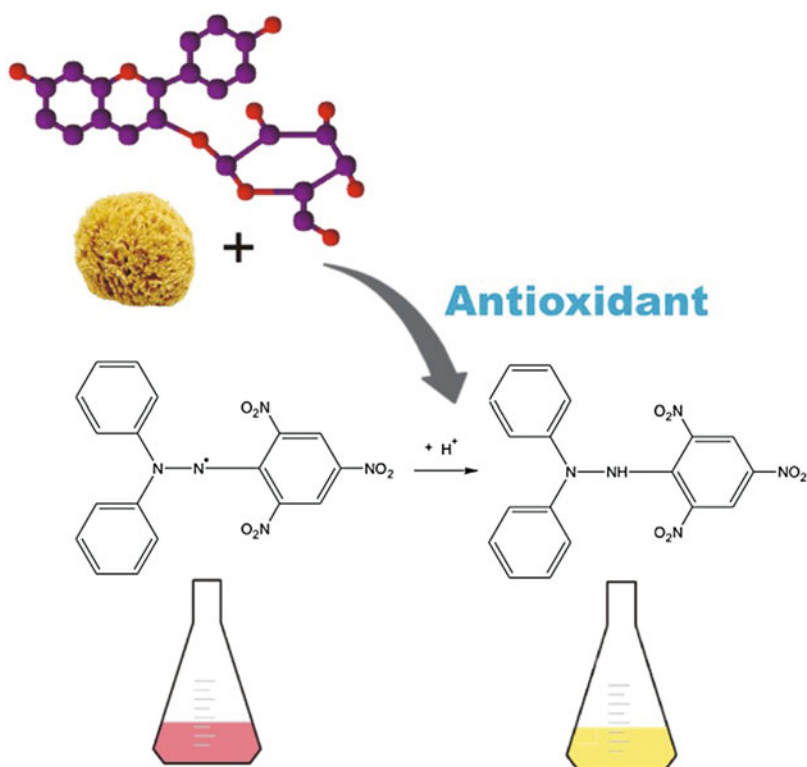


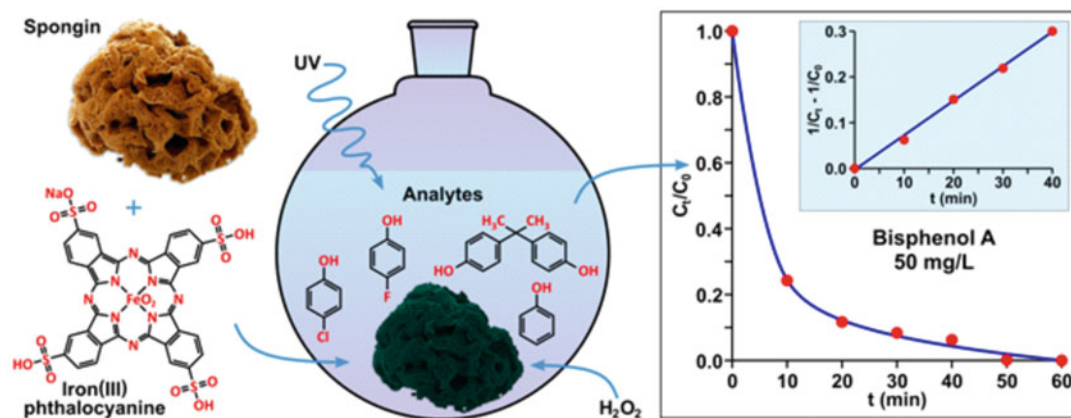
Inorganic, Hybrid, and Functional Pigments, Fig. 10 Different forms of anthocyanins, depending on pH and possible interactions of anthocyanin with spongin at different pH. (Created based on [50] with permission, Copyright 2021: Elsevier Publisher)



Inorganic, Hybrid, and Functional Pigments, Fig. 11 Proposed mechanism of interaction between chlorophyllin and sponginiol. (Created based on Open Access Policy of MDPI Publisher)

Inorganic, Hybrid, and Functional Pigments, Fig. 12 Scheme of antioxidant activity of anthocyanin-sponginiol material. (Created based on [50] with permission, Copyright 2021: Elsevier Publisher)





Inorganic, Hybrid, and Functional Pigments, Fig. 13 Schematic representation of BPA, phenol, chloro- and fluorophenol degradation. (Created based on [54] with permission, Copyright 2021: Elsevier Publisher)

possibility of reusing it. It is important in terms of potential application, among which attention should be given to preparation of wound dressing.

On the other hand, copper or iron phthalocyanine adsorbed onto spongin may exhibit catalytic properties in organic compounds degradation [53, 54]. The catalytic properties of CuPC are retained in combination with spongin and such a material can be utilized in the combined approach of degradation that involves photochemical systems based on photosensitizers absorbing UV light (CuPC), which works as catalyst and external oxidant, H_2O_2 , which form strong oxidizing species, like $OH^\cdot$. Exemplary mechanism of interaction and effectiveness of phenol-based pollutants is presented in Fig. 13.

All of the above mentioned confirms the excellent properties of spongin skeleton as a dye adsorbent, requiring no special preparation of a sponge as well as mild conditions of adsorption process. Moreover, the benefits of spongin, like three-dimensional, anastomosed architecture, their affinity to different structures of dyes, and high sorption capacity make this material a promising candidate for catalyst support. The high sorption capacity is an effect of the structure of skeleton as well as chemical composition of spongin. The adsorption of selected dyes on the surface of marine sponge skeleton may be conducted using relatively simple method with high efficiency and

this process may be controlled by adjusting the basic parameter.

It was proven that properties of functionalized spongin affect the application of this material. Spongin preserves an important property of immobilized dyes and provides chemical and mechanical improvements to the produced material. Hence, novel, functional dye/biopolymer hybrid materials exhibit antioxidant, antimicrobial, and catalytic properties.

Summary

Technologies for the production and application of a wide range of inorganic, hybrid, and multifunctional pigments are closely interrelated and are constantly being developed. In an era in which many branches of industry are undergoing dynamic development, there is a need to improve and obtain new compounds of various kinds, including hybrid, multifunctional pigments, with specific physicochemical and application properties. New trends in the development of pigments relate mainly to:

- The introduction of new pigments or improvement of the properties of existing ones
- Broadening of the assortment and range of uses of pigment composites

- The introduction of nanoparticle-based pigment structures
- Application of naturally occurring supports and colorants in the designing of novel type of functional pigments

The above is confirmed by the numerous R&D projects carried out at scientific centers and in industries, which to date have published more than 86,900 scientific papers related to inorganic pigments (Scopus database 10.01.2021), 69,400 – related to hybrid pigments and more than 205,000 – related to functional pigments.

The main reason for the high level of interest, particularly in the use of nanoparticles, is the possibility of improving the barrier properties of coatings, and thereby also corrosion resistance and mechanical properties. Other drivers of innovation in the field of pigments are environmental awareness, economic pressures, and the constantly changing needs of the industries, which consume these products. Hence the greatest level of development is observed in relation to pigments with special functionality such as anticancer, anti-inflammatory, and antioxidant properties, including hybrid pigments prepared utilizing naturally occurring colorants as well as supports (biopolymers). By using simple procedure of designing new dye-biopolymer pigments increases the possibilities of utilizing those materials and points out perspectives for their further development and practical use. All of these aspects ensure unceasing demand and continued efforts to obtain a wide color range of inorganic, hybrid, and multifunctional pigments, whose importance can be expected to continue to increase.

Cross-References

► Inorganic, Hybrid, and Functional Pigments

Acknowledgments This work was supported by the Polish Ministry of Science and Education. Authors would like to thank Małgorzata Norman, PhD, for her participation and contribution in preparing the updated version of the subsection ‘Colored bioorganic-based hybrid pigments’.

References

1. Garo, D.: Pigments – a growth market. *Europ. Coat. J.* **1**, 10–13 (2012)
2. Endriß, H.: *Inorganic Coloured Pigments Today*. Vincentz Verlag, Hannover (1998)
3. Golgschmidt, A., Streitberger, H.J.: *BASF-Handbuch Lackiertechnik*. Vincentz Verlag, Hannover (2002)
4. Lambournle, R., Strivens, T.A. (eds.): *Paint and Surface Coatings Theory and Practice*, 2nd edn. Woodhead Publishing Ltd., Cambridge (1999)
5. Oyarzún, J.M.: *Pigment Processing Physico-Chemical Principles*. Vincentz Verlag, Hannover (2000)
6. Bauxbaum, G. (ed.): *Industrial Inorganic Pigments*. WILEY-VCH Verlag GmbH, Weinheim (1998)
7. Chen, X., Mao, S.S.: *Titanium dioxide nanomaterials: synthesis, properties, modifications, and applications*. *Chem. Rev.* **107**, 2891–2959 (2007)
8. Moezzi, A., McDonagh, A.M., Cortie, M.B.: Zinc oxide particles: synthesis, properties and application. *Chem. Eng. J.* **185–186**, 1–22 (2012)
9. Klapiszewska, B., Krysztafkiewicz, A., Jesionowski, T.: Highly dispersed green silicate and oxide pigments precipitated from model systems of postgalvanic waste. *Environ. Sci. Technol.* **37**, 4811–4818 (2003)
10. Krysztafkiewicz, A., Klapiszewska, B., Jesionowski, T.: Precipitated green pigments: products of chromate postgalvanic waste utilization. *Environ. Sci. Technol.* **42**, 7482–7488 (2008)
11. Wypych, G.: *Handbook of Fillers*, 3rd edn. ChemTec Publishing, Toronto (2010)
12. Geddes, C.D., Apperson, K., Birch, D.J.S.: New fluorescent quinolinium dyes – applications in nanometer particle sizing. *Dyes Pigments.* **44**, 69–74 (2000)
13. Malinge, J., Allain, C., Brosseau, A., Audebert, P.: White fluorescence from core-shell silica nanoparticles. *Angew. Chem. Int. Ed.* **51**, 8534–8537 (2012)
14. Jesionowski, T., Przybylska, A., Kurc, B., Ciesielczyk, F.: The preparation of pigment composites by adsorption of C.I. Mordant Red 11 and 9-aminoacridine on both unmodified and aminosilane-grafted silica supports. *Dyes Pigments.* **88**, 116–124 (2011)
15. Jesionowski, T., Przybylska, A., Kurc, B., Ciesielczyk, F.: Hybrid pigments preparation via adsorption of C.I. Mordant Red 3 on both unmodified and aminosilane-functionalised silica supports. *Dyes Pigments.* **89**, 127–136 (2011)
16. Siwińska-Stefańska, K., Nowacka, M., Kołodziejczak-Radzimska, A., Jesionowski, T.: Preparation of hybrid pigments via adsorption of selected food dyes onto inorganic oxides based on anatase titanium dioxide. *Dyes Pigments.* **94**, 338–348 (2012)
17. Wu, G., Koliadima, A., Her, Y.-S., Matijević, E.: Adsorption of dyes on nanosize modified silica particles. *J. Colloid Interface Sci.* **195**, 222–228 (1997)
18. Parida, S.K., Mishra, B.K.: Adsorption of styryl pyridinium dyes on polyethyleneglycol-treated silica.

- Colloids Surf. A Physicochem. Eng. Asp. **134**, 249–255 (1998)
19. Binkowski, S., Jesionowski, T., Krysztafkiewicz, A.: Preparation of pigments on modified precipitated silicas. *Dyes Pigments*. **47**, 247–257 (2000)
20. Harris, R.G., Wells, J.D., Johnson, B.B.: Selective adsorption of dyes and other organic molecules to kaolinite and oxides surfaces. *Colloids Surf. A Physicochem. Eng. Asp.* **180**, 131–140 (2001)
21. Neumann, M.G., Gessner, F., Schmitt, C.C., Sartori, R.: Influence of the layer charge and clay particle size on the interactions between the cationic dye methylene blue and clays in an aqueous suspension. *J. Colloid Interface Sci.* **255**, 254–259 (2002)
22. Rytwo, G., Nir, S., Crespin, M., Margulies, L.: Adsorption and interactions of methyl green with montmorillonite and sepiolite. *J. Colloid Interface Sci.* **222**, 12–19 (2000)
23. Hajjaji, M., Kacim, S., Alami, A., Bouadili, A.E., Mauntassir, M.E.: Chemical and mineralogical characterization of a clay taken from the Moroccan Meseta and a study of the interaction between its fine fraction and methylene blue. *Appl. Clay Sci.* **20**, 1–12 (2001)
24. Bujdák, J., Iyi, N., Hrobáriková, J., Fujita, T.: Aggregation and decomposition of a pseudoisocyanine dye in dispersions of layered silicates. *J. Colloid Interface Sci.* **247**, 494–503 (2002)
25. Bujdák, J., Iyi, N.: Visible spectroscopy of cationic dyes in dispersions with reduced-charge montmorillonites. *Clays Clay Minerals*. **50**, 446–454 (2002)
26. Jesionowski, T., Binkowski, S., Krysztafkiewicz, A.: Adsorption of the selected organic dyes on the functionalized surface of precipitated silica via emulsion route. *Dyes Pigments*. **65**, 267–279 (2005)
27. Jesionowski, T., Andrzejewska, A., Krysztafkiewicz, A.: Adsorption of basic dyes from model aqueous solutions onto novel spherical silica support. *Color Technol.* **124**, 165–172 (2008)
28. Winnik, F.M., Keoshkerian, B., Fuller, J.R., Hofstra, P. G.: New water-dispersible silica-based pigments: synthesis and characterization. *Dyes Pigments*. **14**, 101–112 (1990)
29. Faulkner, D.J.: Marine natural products. *Nat. Prod. Rep.* **18**, 1–49 (2001)
30. Munro, M.H.G., Blunt, J.W., Dumdei, E.J., Hickford, S.J.H., Lill, R.E., Li, S., Battershill, C.N., Duckworth, A.R.: The discovery and development of marine compounds with pharmaceutical potential. *J. Biotechnol.* **70**, 15–25 (1999)
31. Lüer, S., Troller, R., Aebi, C.: Antibacterial and anti-inflammatory kinetics of curcumin as a potential antimucositis agent in cancer patients. *Nutr. Cancer*. **64**, 975–981 (2012)
32. Bonarska-Kujawa, D., Cyboran, S., Żyłka, R., Oszmiański, J., Kleszczyńska, H.: Biological activity of blackcurrant extracts (*Ribes nigrum* L.) in relation to erythrocyte membranes. *Biomed Res. Int.* **2014**, 1–14 (2014)
33. Cavalcanti, R.N., Santos, D.T., Meireles, M.A.: Non-thermal stabilization mechanisms of anthocyanins in model and food systems-an overview. *Food Res. Int.* **44**, 499–509 (2011)
34. Shahid, M., Mohammad, F.: Recent advancements in natural dye applications: a review. *J. Clean. Prod.* **53**, 310–331 (2013)
35. Hug, H., Bader, M., Mair, P., Glatzel, T.: Biophotovoltaics: natural pigments in dye-sensitized solar cells. *Appl. Energy*. **115**, 216–225 (2014)
36. Borges, M.E., Tejera, R.L., Díaz, L., Esparza, P., Ibáñez, E.: Natural dyes extraction from cochineal (*Dactylopius coccus*). New extraction methods. *Food Chem.* **132**, 1855–1860 (2012)
37. Paulsen, N., Johnson, R., Coffee, M.: Process for manufacturing chewable dosage forms for drug delivery and products thereof. U.S. 20070128251 A1 patent, (2007)
38. Park, K.H., Kim, T.Y., Park, J.Y., Jin, E.M., Yim, S.H., Choi, D.Y., Lee, J.W.: Adsorption characteristics of gardenia yellow as natural photosensitizer for dye-sensitized solar cells. *Dyes Pigments*. **96**, 595–601 (2013)
39. Friedman, M.: Antibacterial, antiviral, and antifungal properties of wines and winery byproducts in relation to their flavonoid content. *J. Agric. Food Chem.* **62**, 6025–6042 (2014)
40. Thomasset, S., Berry, D.P., Cai, H., West, K., Marczylo, T.H., Marsden, D., Brown, K., Dennison, A., Garcea, G., Miller, A., Hemingway, D., Steward, W.P., Gescher, A.J.: Pilot study of oral anthocyanins for colorectal cancer chemoprevention. *Cancer Prev. Res.* **2**, 625–633 (2009)
41. Hayatsu, H., Sugiyama, C., Arimoto-Kobayashi, S., Negishi, T.: Porphyrins as possible preventers of heterocyclic amine carcinogenesis. *Cancer Lett.* **143**, 185–187 (1999)
42. Aozasa, O., Tetsumi, T., Ohta, S., Nakao, T., Miyata, H., Nomura, T.: Fecal excretion of dioxin in mice enhanced by intake of dietary fiber bearing chlorophyllin. *Bull. Environ. Contam. Toxicol.* **70**, 359–366 (2003)
43. Yamazaki, H., Fujieda, M., Togashi, M., Saito, T., Preti, G., Cashman, J.R., Kamataki, T.: Effects of the dietary supplements, activated charcoal and copper chlorophyllin, on urinary excretion of trimethylamine in Japanese trimethylaminuria patients. *Life Sci.* **74**, 2739–2747 (2004)
44. Falabella, A.: Debridement and wound bed preparation. *Dermatol. Ther.* **19**, 317–325 (2006)
45. Wohrle, D., Schnurpfeil, G., Makarov, S.G., Kazarin, A., Suvorova, O.N.: Practical applications of phthalocyanines – from dyes and pigments to materials for optical, electronic and photo-electronic devices. *Macroheterocycles*. **5**, 191–202 (2012)
46. Yao, Y., Chen, W., Lu, S., Zhao, B.: Binuclear metallophthalocyanine supported on treated silk fibres as a novel air-purifying material. *Dyes Pigments*. **73**, 217–223 (2007)
47. Rismayani, S., Fukushima, M., Sawada, A., Ichikawa, H., Tatsumi, K.: Effects of peat humic acids on the catalytic oxidation of pentachlorophenol using

- metalloporphyrins and metallophthalocyanines. *J. Mol. Sci. A Chem.* **217**, 13–19 (2004)
48. Wei, Z., Cao, Y., Ma, W., Wang, C., Xu, W., Guo, X., Hu, W., Zhu, D.: Langmuir–Blogett monolayer transistors of copper phthalocyanine. *Appl. Phys. Lett.* **95**, 33304 (2009)
 49. Bialek, B., Kim, I.G., Lee, J.I.: Electronic structure of copper phthalocyanine monolayer: a first-principles study. *Thin Solid Films.* **436**, 107–114 (2003)
 50. Norman, M., Bartczak, P., Zdarta, J., Ehrlich, H., Jesionowski, T.: Anthocyanin dye conjugated with *Hippospongia communis* marine demosponge skeleton and its antiradical activity. *Dyes Pigments.* **134**, 541–552 (2016)
 51. Al-Degs, Y.S., El-Barghouthi, M.I., El-Sheikh, A.H., Walker, G.M.: Effect of solution pH, ionic strength, and temperature on adsorption behavior of reactive dyes on activated carbon. *Dyes Pigments.* **77**, 16–23 (2008)
 52. Gómez, J.M., Galán, J., Rodríguez, A., Walker, G.M.: Dye adsorption onto mesoporous materials: pH influence, kinetics and equilibrium in buffered and saline media. *J. Environ. Manag.* **146**, 355–361 (2014)
 53. Norman, M., Bartczak, P., Zdarta, J., Tomala, W., Żurańska, B., Dobrowolska, A., Piasecki, A., Czaczyk, K., Ehrlich, H., Jesionowski, T.: Sodium copper chlorophyllin immobilization onto *Hippospongia communis* marine demosponge skeleton and its antibacterial activity. *Int. J. Mol. Sci.* **17**, 1–17 (2016)
 54. Norman, M., Żółtowska-Aksamitowska, S., Zgoła-Grześkowiak, A., Ehrlich, H., Jesionowski, T.: Iron (III) phthalocyanine supported on a spongin scaffold as an advanced photocatalyst in a highly efficient removal process of halophenols and bisphenol A. *J. Hazard. Mater.* **347**, 78–88 (2018)

Instrument: Colorimeter

Shao-Tang Hung

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Tristimulus colorimeter

Definition

There are two kinds of modern colorimeters: one that includes and one that excludes a light source for measuring the surface or a self-luminous

stimulus respectively. Both types of colorimeter measure International Commission on Illumination (CIE) tristimulus values under particular illuminant and observer conditions, e.g., D50/2°. The instruments are applied to measure a test stimulus displayed on a monitor, printer or a light stimulus. CIE International Lighting Vocabulary [1] defines the colorimeter as an instrument for measuring colorimetric quantities, such as the tristimulus values of a color stimulus.

Overview

The colorimeter measures XYZ values in a luminance unit. This overview is divided into four sections: optical structure, measurement geometry, spectral responsivity and measuring uncertainty, and applications [1] (► [Photodetectors](#)).

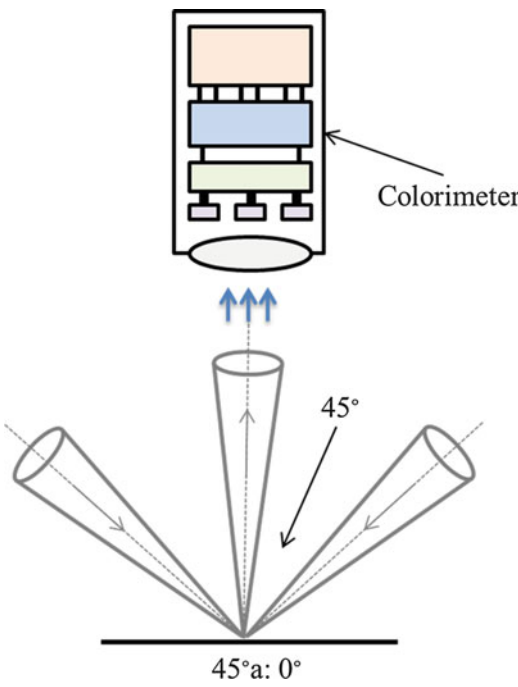
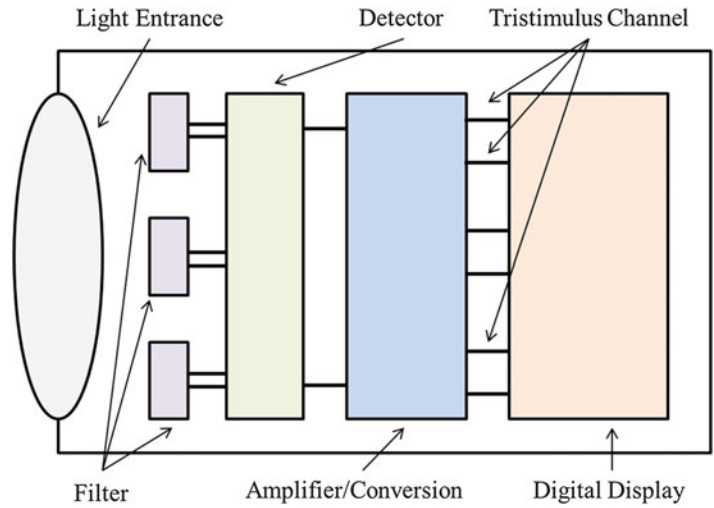
Optical Structure

The structure of the colorimeter can be divided into two parts: the measuring head and the digital display. Figure 1 shows the schematic structure of a colorimeter. The measuring head contains tristimulus filters, detectors, and spatial corrections of the detector response. There are filters in front of the detectors to filter the input light, which is from the entrance window. The filters mimic the CIE color-matching functions. The parallel photodetectors sense the filtered light in the electronic signal ► [Photodetectors](#). The multiplex amplifier can integrally form the signal to feed via output channels. Finally, the XYZ tristimulus is reported on a digital display [2].

Measuring Geometry

There are three factors affecting the measurement results: the limiting aperture, the acceptance area, and the optical axis of the colorimeter head. The limiting aperture affects the quantity of the incoming light. The acceptance area evaluates the incident radiation to improve accuracy. The optical axis is the middle point of the measured surface or light source, which is related to the light entrance of the measuring head [3]. Figure 2 reveals 0°a/45° measuring geometry; more details are shown under (► [Photometer](#)).

Instrument: Colorimeter,
Fig. 1 Schematic structure
of a colorimeter [3]



Instrument: Colorimeter, Fig. 2 0°a/45° measuring
geometry

**Spectral Responsivity and Measuring
Correction**

The measured spectrum is reported across a visible spectrum of 380–780 nm. Because the color-matching functions (CMF) $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ approximate the three human spectral responsivities, the real relative spectral responsivities (RSR) $s_{x,r}(\lambda)$, s_y ,

$s_z(\lambda)$, $s_{z,r}(\lambda)$ show some deviation. They do not match the color-matching functions exactly.

The mismatch correction of the spectrum must be considered; many of the corrected factors are used to evaluate the mismatch and to correct these deviations. For example, there are three types of factors corrected by comparing actual and target functions: the spectral mismatch, ultraviolet response ($f_{UV,i}$), and infrared response ($f_{IR,i}$). The corrected indexes are also used for the correction of the spectral transmittance factor ($\bar{\tau}(\lambda)$).

Applications

The colorimeter can measure XYZ tristimulus values in the illuminant industry, in paper reflectance, and in digital imaging. All of the measuring data, which calculate the color tristimulus to compare the test with reference values, following the CIE CMF, are employed, the user demand to determine the visual angle, and choose the corresponding CMF data (2°/10°) [2]. Moreover, the reference illuminant is also considered when the colorimeter calculates the XYZ tristimulus [4].

Most important for the colorimeter is that it is used to assist in calibrating color devices. For a printer, the colorimeter measures the different color patches that are printed on the paper. The measurement results are compared with the reference printer data, and the color difference is calculated to ensure colorimetric consistency.

Applications in the lighting industries can measure XYZ tristimulus in units of luminance (cd/m^2). The luminance can be represented clearly to correct the illuminant. Furthermore, the hand-held colorimeter can be conveniently used outdoors to measure the correlated color temperature (T_{cp}) and luminance.

Cross-References

► [Photodetector](#)

References

1. International Commission on Illumination: ILV: International Lighting Vocabulary. CIE S017/E:2011, Vienna (2011)
2. International Commission on Illumination: Methods for Characterising Tristimulus Colorimeters for Measuring The Colour of Light. CIE 179:2007, Vienna (2011)
3. Ohta, N., Robertson, A.: Colorimetry: Fundamentals and Applications. Wiley, Hoboken (2006)
4. X-Rite: A Guide to Understanding Color Communication. Grand Rapids (2007)

Instrument: Photometer

Shao-Tang Hung
Center for Measurement Standards,
Industrial Technology Research Institute,
Hsinchu, Taiwan

Synonyms

[Luminous intensity meter](#)

Instrument: Photometer,
Instrument: Photometer,
Fig. 1 Luminous intensity
measurement [2]

Definition

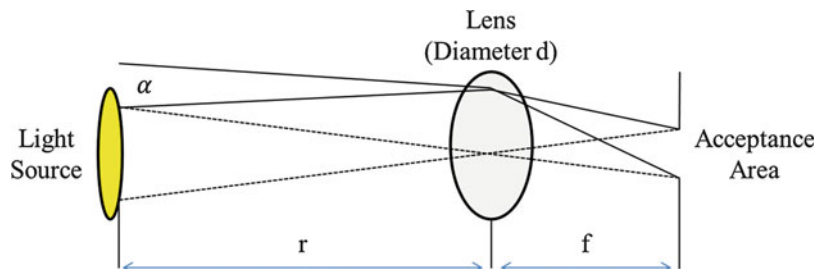
The photometer is an instrument which can be used to measure photometric quantities such as luminance, illuminance, etc. The parameters are derived from means of law of additives and CIE spectral luminous efficiency functions. The instruments usually transport the measured data either digitally or analogly to display on the screen. CIE had defined the international vocabulary of the photometer: instrument for measuring photometric quantities [1].

Overview

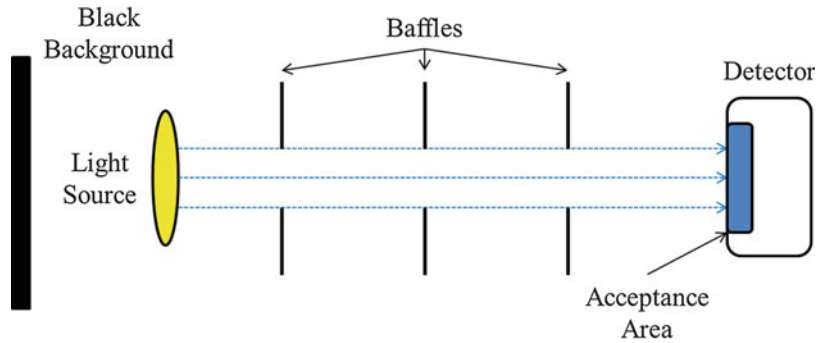
The photometer is a universal vocabulary which has plenty of types. According to the applications, the types can be divided to luminance meter, illuminance meter, etc. The luminous-intensity meter is a conventional photometer, which can measure directly luminous intensity (candela, cd). This article will focus on the luminous-intensity meter and is divided into three sections to introduce luminous-intensity meter structure and measuring theories, measuring fields and range, and applications and comparison [1–3].

Luminous-Intensity Meter Structure and Measuring Theories

The luminous-intensity meter consists mainly of detector unit. Following the measuring theories, the spectral weighting or the spectral dispersion device is arranged in front of the detector to separate the light. The lenses and apertures assist light transportation for luminous intensity measurement, see Fig. 1. The use of additional lenses can change the field of the focus to make it more



Instrument: Photometer,
Instrument: Photometer,
Fig. 2 The perfect
 conditions for luminous-
 intensity meter
 measurement [2]



flexible [4]. Luminous intensity can be ensured by measuring theories.

The data is measured by a detector which is included in the measuring detector that is transformed in terms of the digital signal, pass-through amplifier, and digital processes. If users must calculate other illuminant parameters such as luminance, illuminance, etc., the data can be outputted to the computer for further analysis.

Acceptance Area

Following the CIE definition, the acceptance area is the area of the luminous-intensity meter detector which is receiving and directionally evaluating the incident light. The acceptance area affects directly luminous intensity. In manufacturing processes of luminous-intensity meter, the measuring field must be defined accurately.

The measurement results are affected certainly by the environmental condition such as the background. The entire scene which includes light is also evaluated before the light input to the luminous-intensity meter detector. In the metrologically perfect conditions, the explanation is presented in Fig. 2. The light source is installed in front of the black background. The stray light is sheltered by baffles; the detector can receive the purer light that is not interfered with by the other nonstationary light [2].

Applications

The SI unit of luminous intensity is the candela (cd). The luminous-intensity meter can be used to measure light source in a particular direction per unit solid angle. The luminous-intensity meter is an adaptive tool to calibrate the optical

instruments. The sensitive detector is also applied to measure flicker or nonstationary optical wave. Furthermore, the other types of photometers such as luminance meter and illuminance meter can measure environmental variety. For example, luminance and illuminance are measured continuously for realizing LED light and overall light respectively. The correlation also can be found.

References

1. International Commission on Illumination: ILV: International Lighting Vocabulary, CIE S017/E:2011, Vienna (2011)
2. International Commission on Illumination: The Measurement of Absolute Luminous Intensity Distributions, CIE 70–1987, Vienna (2011)
3. International Commission on Illumination: Methods of Characterizing Illuminance Meters and Luminance Meters, CIE 69–1987, Vienna (2011)
4. International Commission on Illumination: Measurement of Luminous Flux, CIE 84–1989, Vienna (2011)

Instrument: Spectrophotometer

Shao-Tang Hung

Center for Measurement Standards, Industrial
 Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Spectroreflectometer \(for reflectance measurement\)](#)

Definition

The spectrophotometer is an instrument for measuring the reflectance or transmittance quantities of an object, including specular and diffuse reflectance and regular and diffuse transmittance. On request, the instrument can measure the above quantities at different viewing/illumination geometry. The International Commission on Illumination (CIE) defines the spectrophotometer in its International Lighting Vocabulary [1] as an instrument for measuring the ratio of two values of a radiometric quantity at the same wavelength.

Overview

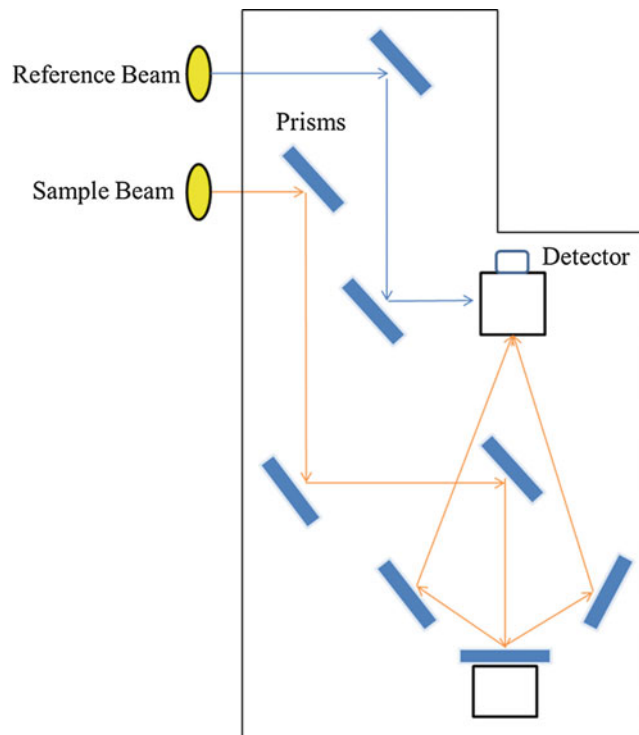
The spectrophotometer is used to measure spectral reflectance or transmittance across a visible spectrum. The structure consists mainly of a light source, a system of optical transmission, and a detector. Furthermore, the viewing geometries

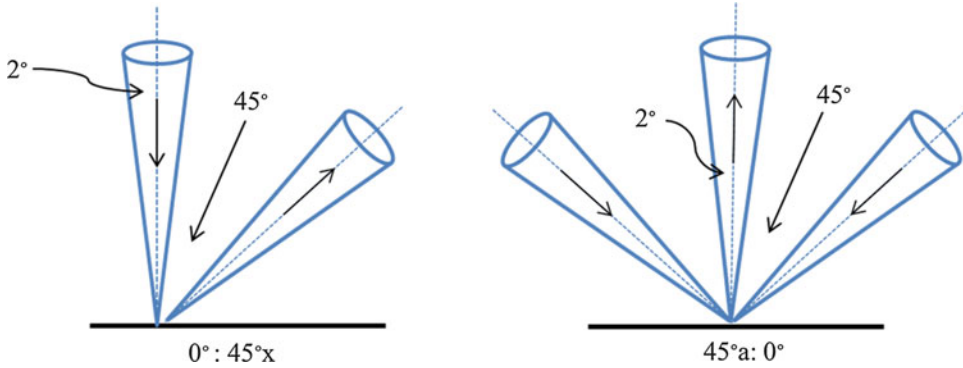
also affect the measurement results. The instrument is composed of the following three parts: structure, viewing/illumination geometries, and application [1, 2].

Structure

The spectrophotometer consists of multiple structures to carry out light transmission and sample measurement. First of all, the measurement condition divides the wavelength into three sections: UV, visible, and IR sections. The deuterium lamp is used to measure the UV section. The visible and IR sections are measured by the halogen lamp. Second, the beam is divided into the reference beam and the sample beam. The reference beam and the sample beam are used to monitor stability and measure the sample, respectively. Finally, the two beams are transmitted by the multiple prisms and focused on the detector. The optical transmission is revealed in Fig. 1.

Instrument:
Spectrophotometer,
Fig. 1 The optical
transmission of a
spectrophotometer [3]





Instrument: Spectrophotometer, Fig. 2 Standard geometries for spectral reflectance [1]

Viewing/Illumination Geometries

The geometries for measuring reflectance depend on radiation field and angle; several standardized geometries are defined by the CIE. For color measurement, some of the standard geometries are shown in Fig. 2. The spectrophotometer commonly uses bidirectional geometry $0^\circ:45^\circ$ to calibrate the color in the National Measurement Laboratory R.O.C.

The Field of View

The tristimulus values are calculated from spectral reflectance, the factors are affected by the field of view ($2^\circ/10^\circ$). According to the CIE definition, the field of view can be decided by a range; if more than 4° , the field of view is 10° , and if less than 4° , the field of view is 2° . The field of view is one of the important elements for evaluating color. According to the field of view, the color-matching function (CMF) is divided into two types: the CIE 1931 2° Standard Observer and the CIE 1964 10° Standard Observer. The two types of CMF are used to fit the different measurement conditions.

Applications

For color measurement, the spectrophotometer measures a variety of surface colors, such as

textiles, coatings, and plastics. The results are reported as spectral reflectance or transmittance across the visible spectrum. The spectrophotometer is the appropriate device for measuring color data, and provides the correct color information for the manufacturing of textile products, printer products, and paint and material products to solve color difficulties.

References

1. Schanda, J. (ed.): *Colorimetry: Understanding the CIE System*. Wiley, Hoboken (2007)
2. International Commission on Illumination: *ILV: International Lighting Vocabulary*. CIE S017/E: 2011, Vienna (2011)
3. Industrial Technology Research Institute Center for Measurement Standards: *Spectrophotometer MSVP*. Hsinchu (2009)

Integrating Sphere

Cheng-Hsien Chen
Center for Measurement Standards,
Industrial Technology Research Institute,
Hsinchu, Taiwan

Synonyms

Ulbricht sphere

Definition

An integrating sphere is a hollow spherical cavity with a diffuse-reflecting material inside the sphere that has a spatially uniform surface, and a flat profile of spectral power distribution. The structure of integrating spheres generally consists of a baffle, different apertures, a light source, and a detector. The ideal integrating sphere has the following characteristics:

- (i) Small detector and external port.
- (ii) Miniature devices (light source, baffle, and mount) inside the sphere, which cannot affect optical radiation after its first reflection.
- (iii) The surface of the sphere is a homogeneous Lambertian reflector, with reflectance independent of the wavelength.

In color applications, an integrating sphere is used to produce a uniform light source from a luminaire or for measuring the transmittance or reflectance of a surface in a particular viewing geometry [1, 2].

Overview

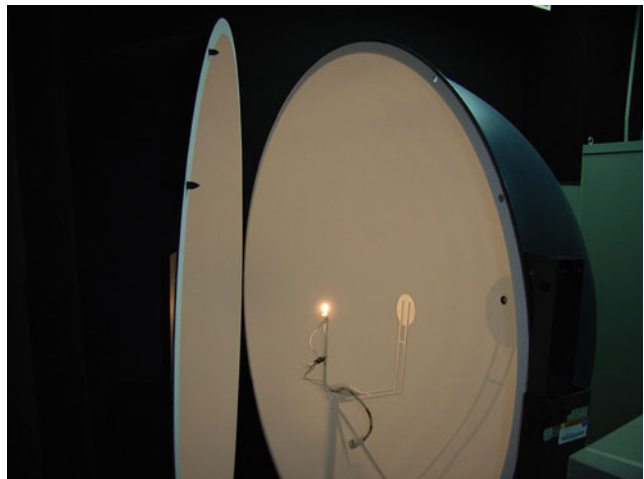
As mentioned earlier, one application is for the measurement of the color properties of a

luminaire. In general, the substitution method is utilized to calculate the color properties of a test and a standard lamp, or transfer standard. The standard lamp provides the spectral flux, which can calculate the colorimetry and photometry quantities. Two types of detector are used on the sphere: a broadband $V(\lambda)$ -corrected photometer, and a narrow-band spectroradiometer [3]. The spectral mismatch error of the calibrated photometer should be corrected, according to the technical report of the International Commission of Illumination (CIE) [4]. The spectral response of the photometer is easy to apply to the measurement of the luminous flux of a test lamp. The spectroradiometer may be applied for spectral power distribution and colorimetry of test lamp measurement (Fig. 1).

When measuring surface properties, an integrating sphere is used as a uniform lighting source or a diffuser of the incident light. For the surface property measurement system, the integrating sphere is set to diffuse the reference beam and geometry structure to analyze the characteristics of the surface. According to various standard geometries, the integrating sphere provides various conditions to realize different optical properties (photometry, colorimetry, haze, gloss, clarity, shading, etc.) of the surface colorimetry and photometry measurement [5, 6].

The integrating sphere method requires a standard lamp, and the substitution method is often

Integrating Sphere,
Fig. 1 An image of an integrating sphere. The baffle inside the integrating sphere is designed to directly avoid the light incident on the detector port. The shadow from the baffle must cover the detector



utilized to calculate the total luminous flux of the test lamp. There are some important components inside the integrating sphere to consider. One such device is the baffle. The size of the baffle, used to prevent incident light from reaching the detector, depends upon both the size of the detector, and the distance between the lamp and the detector. A second device to consider is an auxiliary lamp. In practice, the shape of the test lamp is different from that of the standard lamp. The auxiliary lamp can provide the absorption correction function between the standard lamp and the test lamp.

The equation of luminous flux Φ_t is

$$\Phi_t = \frac{I_t}{I_s} \cdot \frac{A_s}{A_t} \cdot \Phi_s \quad (1)$$

where I_t is the signal of the test lamp, I_s is the signal of the standard lamp, A_s is the signal of the auxiliary lamp with the standard lamp, and A_t is the signal of the auxiliary lamp with the test lamp.

The photodetector shows some mismatches between the spectral luminous efficiency function and the spectral response of the detector and lamp source relation. The nonuniformity of the integrating sphere causes the directional light source to contribute measurement errors inside the sphere. The baffle size and different geometry inside the sphere provide varying spatial response distribution. Therefore, the equation of luminous flux Φ_t is given by

$$\Phi_t = \frac{I_t}{I_s} \cdot \frac{A_s}{A_t} \cdot ccf \cdot scf \cdot \Phi_s, \quad (2)$$

where ccf is the spectral mismatch correction factor, and scf is the spatial response distribution function (SRDF) of the integrating sphere [7, 8].

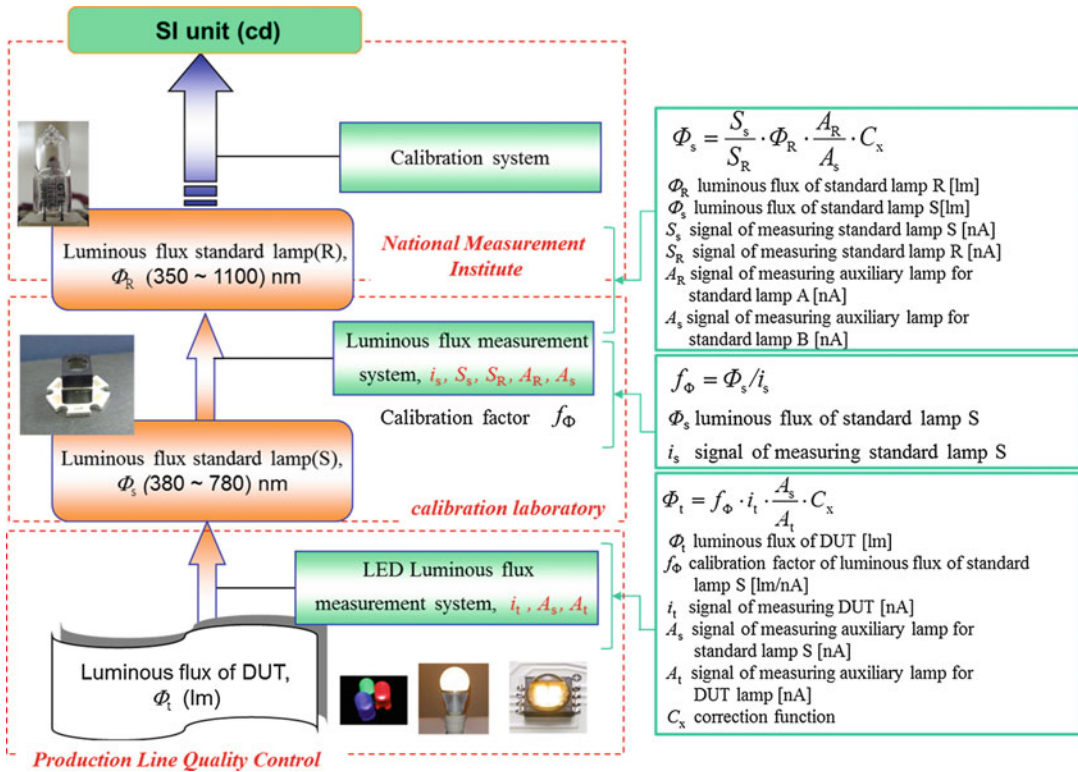
A good measurement system of an accreditation laboratory should have a suitable traceability chain. The metrological traceability chain, according to the International Standards Organization (ISO) guide 99 is the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. Therefore, the product

line quality control of manufacturing needs a transfer standard from the calibration laboratory. The calibration laboratory uses the reference standard trace to the National Measurement Institute (NMI), and then the trace to the SI unit shown in Fig. 2.

The stable and repeatable transfer standards, which have similar characteristics to the test lamp, are needed for the measurement system to transfer photometric and colorimetric units from the NMI or an accreditation laboratory. In practice, the transfer standard is made by a light-emitting diode (LED) lamp whose optical characteristics are similar to those of the test lamp and are measured using the same measurement instrument. For the calibration laboratory, the luminous flux measurement system transfers the standard from the reference standard to the LED transfer standard. The NMI transfers the standard, which is made by a stable, repeatable, and broadband spectrum lamp, such as a tungsten or halogen lamp, to the calibration laboratory.

For the luminous flux measurement of lighting, an integrating sphere is sometimes used together with calibration with a standard lamp that has similar characteristics and by using a comparison method to transfer luminous flux quantity to the measurement instrument. Measurement of LED lighting involves two kinds of measurement geometries. First, the test sample is mounted on the integrating sphere. Second, the test lamp is located in the center of the sphere. Why do the test lamp and standard lamp have similar characteristics? Because the measurement result is achieved by comparing the input and output signals. If we consider the spatial nonuniformity of integrating, by rotating a scanning beam, we will find the nonperfect output of the integrating sphere. If we use the narrow beam light source to scan inside the sphere, we measure the different signal from the detector in the different scanning region shown in Fig. 3.

The 2π geometry would have two zones, and zone 2 would have a drift signal as the large scanning beam angle. This means that when using a wide diverging angle lamp such as a halogen lamp or a diffused lamp such as a standard lamp, and the narrow diverging angle of the



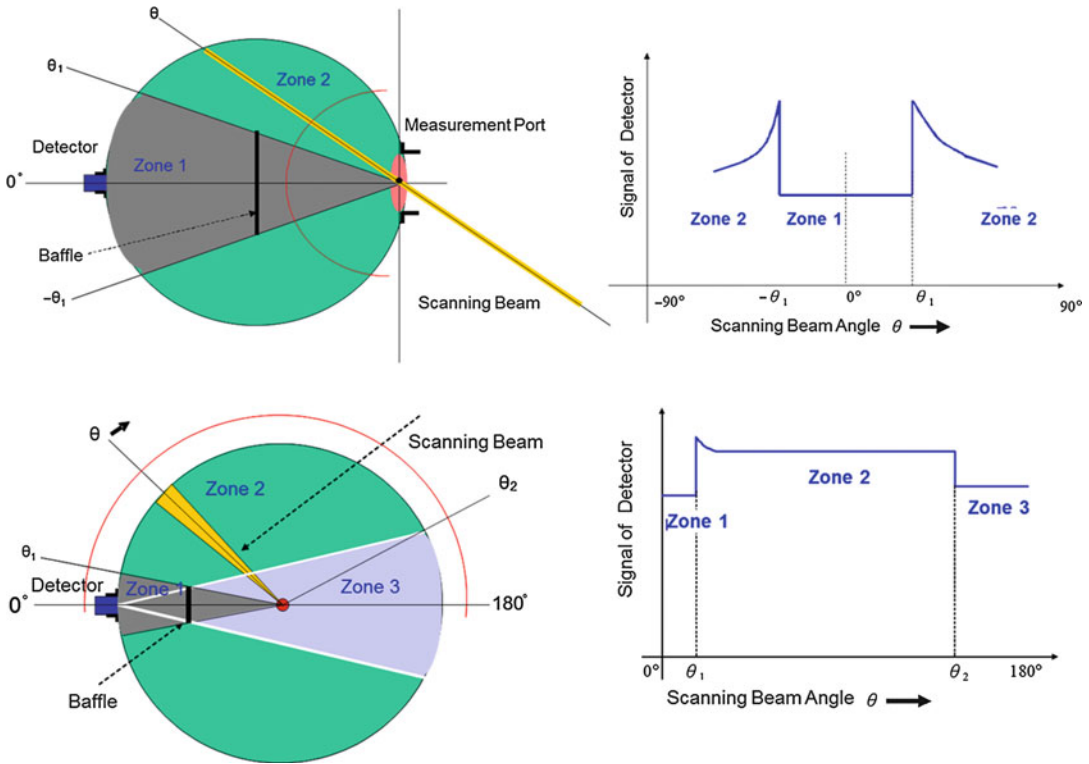
Integrating Sphere, Fig. 2 The metrological traceability chain of a luminous flux standard lamp calibration system

test lamp such as some 5-mm LEDs, there would be more measurement uncertainty for calibration, because the large size of the integrating sphere and perfect Lambertian surface were good for measuring the luminous flux of the LED. For the 4π geometry, zone 1 is the baffle region with the scanning beam near the edge of the baffle, and we would measure a higher signal from the detector. Some light is directly incident to the detector because of scattering. If we use a large sphere and a baffle and mount of smaller geometry, a large area of the flat region would be zone 2. Zone 3 is the shadow region made from the baffle and detector by the scanning beam.

Most important is the light distribution between the standard lamp and the test lamp. If the light distribution is different, the spatial non-uniformity of the integrating sphere should be considered. The nonuniformity of the integrating sphere causes the directional light source to make a measurement error inside the sphere. The

baffle size and different geometry inside the sphere provide a different spatial response distribution.

Ideally, the light distribution of the device under test (DUT) lamp and the standard lamp should be similar. If not, the spatial distribution mismatch correction is applied complexly. If measurement of a directional light source is required, for which the light distribution is different from the standard lamp, the spatial response distribution function (SRDF) should be considered. The SRDF is relative to the baffle size, the DUT lamp, and the geometry inside the sphere. In practice, the point source in the center of the sphere should be used and the light distribution of the DUT measured to calculate the SRDF. In addition, the same method of evaluating the SRDF to measure the directional light source in the center of the integrating sphere can be used. Overall, by knowing the limitation of the integrating sphere, the directional light source can be precisely



Integrating Sphere, Fig. 3 Spatial nonuniformity of the integrating sphere

measured, and the measurement uncertainty can be reduced [9].

Cross-References

- [Spectral Power Distribution](#)
- [Spectroradiometer](#)
- [Standard Lamp](#)

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. Gigahertz-Optik website. <http://www.light-measurement.com/ideal-integrating-sphere/> (2014)
3. Illuminating Engineering Society of North America.: Electrical and Photometric Measurements of Solid-State Lighting Products. IES LM-79-08, New York (2008)
4. International Commission on Illumination.: Measurement of LED. CIE 127:2007, Vienna (2007)
5. International Commission on Illumination.: Measurement of Luminous Flux. CIE 84:1989, Vienna (1989)
6. International Commission on Illumination.: The Basis of Physical Photometry. CIE 18.2:1983, Vienna (1983)
7. Ohno, Y.: Integrating sphere simulation: application to total flux scale realization. *Appl. Optics*. **33**(13), 2637–2647 (1994)
8. Ohno, Y., Daubach, R.: Integrating sphere simulation on spatial nonuniformity errors in luminous flux measurement. *J. IES*. **30**(1), 105–115 (2001)
9. Chen, C.H., Pong, B.J., Jiaan, Y.D., Lin, H.L.: Lamp orientation dependence of an integrating sphere response for directional light source in luminous flux measurement. *NCSLI Measure J. Meas. Sci.* **8**(2), 42–47 (2013)

Integrative Lighting

- [Human-Centric Lighting](#)

Interactions of Color and Light

Ulrich Bachmann¹ and Ralf Michel²

¹Institute of Colour and Light, Zurich, Switzerland

²Institute Integrative Design | Master Studio Design, University of Applied Sciences FHNW Basel, Basel, Switzerland

Synonyms

Color and light effects; Color and light interdependence; Influences between light and color

Definition

Color cannot be observed without light, and without material artifacts light is not visible – thus, one can simplify and say: no color exists without light and no light could be present without color. Color is a visual sensation that depends, among other factors, on stimulation of the visual system by radiation within the visible range – light – either emitted by a source or transmitted or reflected by *pigmented* objects. Indefectibly, light has color; there is no colorless light (light can be white, but white is also a color sensation). The chromatic composition of incident light affects the perceived color of pigmented surfaces. Color and light are like both sides of the same coin; they cannot be separated. For the creative disciplines involving design (design, *interior design*, *scenography*, architecture, etc.), the use of color and light and the knowledge about how one interacts and influences each other is a fundamental issue.

Interactions of Color and Light in the Context of Creative Training and Practices

Initial Position

In creative training and practices, the determining aspects of the interactions of color and light are

generally not represented in a systematized manner. The present authors have responded to this shortcoming in collaboration with Florian Bachmann, Marcus Pericin [1], and others by examining the relevance of the interaction of color and light and by describing the aspects of it which are most important to creative training and practices. In addition, they use different approaches to exemplify the makings of a new teaching model.

Color cannot be observed without light, and without material artifacts light is not visible. Thus, one can simplify and say: no color without light and no light without color. For the creative disciplines, the use of color is one of the fundamental dimensions. And yet, despite the reciprocal interdependence of color and light, there has as yet been no comprehensive and systematizing concept for theory and creative practice.

Research on the Interactions of Color and Light in the Creative Disciplines

This entry is based on three research projects on the interactions of color and light which build upon each other's findings: ColorLightLab, LED-ColorLab, and Colour and Light – Materials for a Theory of Colour and Light [2].

The research starts from the premise that design systems using color and light which are employed in design practice (and by analogy in *interior design*, *lighting design*, etc., as well) do not relate to each other. This despite the fact that material color cannot be perceived without light and that light cannot be seen in the absence of material artifacts. This shortcoming is reflected in the practice and methods of creative teaching. There is no theory which combines color and light in a systematic manner. The aim of the research was to pursue a practically oriented investigation of the principles involved in the interaction of color and light in terms of the requirements imposed by design and, in addition, to express research findings in practically relevant materials towards a color-light theory.

ColorLightLab

This project investigated the principles of the interactions of light and material color. The research was carried out using the methods of artistic research and research through design [3–5] and culminated in a presentation of the phenomena under investigation and the publication *Colours Between Light and Darkness* [6].

LED-ColorLab

The focus was on an investigation of the interaction of dynamic light and colored surfaces in a spatial context. Here the interactions of dynamic light in the color spectrum were systematically observed, as were those in the spectrum of white light. The findings were presented in a laboratory display and evaluated in surveys with experts (architects, *interior designers*, designers, *lighting designers*, etc.) and with consumers. Approaches from the realm of research through design were combined with survey methods.

Colour and Light – Materials for a Theory of Colour and Light

Here the focus was on conveying the practical relevance of the previous findings. Materials for a theory of color and light (installations, digital and physical tools) were developed in an experimental context and evaluated by experts as well as school and university students (methods used were research through design as well as qualitative and quantitative surveys). These materials were summarized in a multimedia publication [7] which comprises: description of the installations, contextualization of practice, theory and research, compendium of a DVD with videos about the installations, a specifically developed software, and a relaunched website setting attitudes to the interaction of color and light in the context of historical systems of color.

The experiments and empirical investigations were set up on the basis of the following research questions:

- How is the *perception of color* on surfaces and in spaces affected by lighting?

- What effect does the physical composition of surfaces, materials, and spaces have on the perception of color under different lighting?
- What interactions occur between dynamic light and colored surfaces?
- How can one arrive at a comprehensible description of the interaction of dynamic light and colored surfaces?
- What are the requirements to be met by materials and instruments used in a theory of color and light?

These questions derive from experiences gained in design practice and theory.

Summary of Investigations

First, models and prototypes were built, and several test measurements were carried out to compare the differentiation of various surfaces and various light sources; then a measuring system was constructed, where calibrated light falls on two precisely defined and identical surfaces in such a way that the remitting light can be measured by means of sensors. Thus, a kind of color-sampling booth (*LED-ColorCase*) was produced, enabling surface colors under different light sources to be visually compared and methodologically described. The *LED-ColorCase* makes it possible to record and evaluate *hue* shifts visually and in terms of spectral analysis. What is striking is the powerful influence of the light sources used on the difference of the resultant color effects. It was also possible to verify the possibility of calculating and predicting the resultant color impressions.

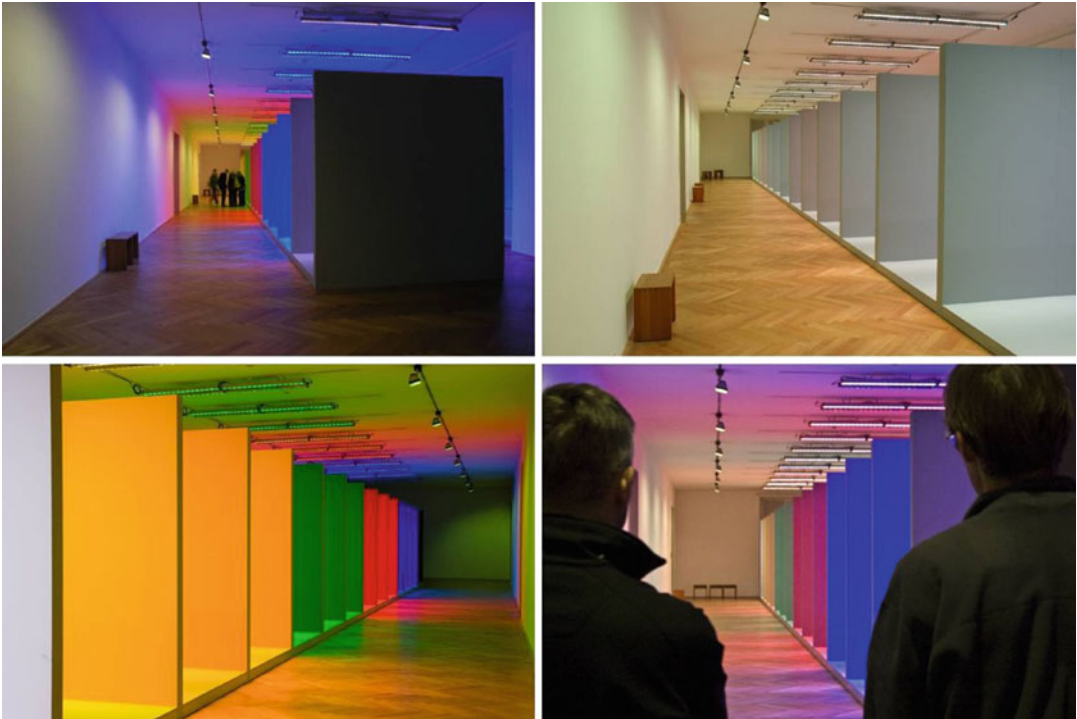
The experiments and laboratory investigations in the *LED-ColorLab* project showed that hue shifts develop a far greater differential under different light sources than was previously supposed and that they are clearly perceptible. In general terms one can say that the effect of and alteration caused by *hue*, *saturation*, and *brightness* is underestimated and generally not sufficiently taken into account in color and light design. In an ideal situation, it should be possible to give such alterations primacy during planning and to move towards predicting them.

Summary of Surveys and Empirical Aspect

The empirical aspect of the *LED-ColorLab* research project resulted in clear findings in terms of practical applicability. Surveys carried out during the LED – Staging Color and Light [8] exhibition (a public laboratory situation and part of the *LED-ColorLab* research project) showed that one can recognize the importance of the interactions of dynamic light (across the whole color spectrum but also dynamically in the white light spectrum from cool white to warm white): experts in the fields of architecture, *interior design*, design, and *lighting design* rate this potential highly in terms of their own creative practice. At the same time, the experts who were questioned demonstrated a lack of methodological and technical knowledge. End users questioned in the *LED-ColorLab* installations

use recent lighting technology either as mood lighting or task lighting (this relates to *LED*). They rate highly the potential for integrated use of color and light in their personal environment and their place of work. Regarding the effects of the color shifts described above, acceptance is almost at 90% overall.

Similar results are seen in analysis of the responses to the research project *Colour and Light – Materials for a Theory of Colour and Light*, which took as its focus the development and evaluation. In the course of in-depth interviews and on the basis of surveys in the display of materials, 71% of the experts questioned acknowledged not only the quality of the materials produced but also the underlying approach towards a theory of color and light for creative practice.



Interactions of Color and Light, Fig. 1 Color-light keyboard installation; project: staging *LED-Color*. It is a 25-m-long spatial installation, which consists of 15 large-scale wall panels painted on one side in the same gray and on the other with different, highly pigmented monochrome colors: white, gray, green, red, ultramarine, and black. The surfaces are illuminated with changing colored LED

lighting according to an animated score of color and light. Because the installation can be entered on foot, the dynamic interactions of color and light can be directly experienced in a spatial context. In this way the installation points towards the great scenic and dramaturgical potential of such compositions of color and light for design and the performing arts

Implementing Knowledge as Mediation Tools and Materials

The principles of interactions of dynamic light and colored surface were laid out in the multimedia publication *Colour and Light: Materials for a Theory of Colour and Light* [7]. Here the aspect of direct sensual perception is given especial prominence. The authors are convinced that a theory of light and color has a particularly strong influence on creative practice if it exploits direct sensual perception of the phenomena of the interactions of color and light. In this the authors are following, among other influences, the concept of the German philosopher Gernot Böhme who writes, using the argument of an aesthetic (a general theory of perception) about the effects of atmospheres: “the primary subject of sensual perception is not the things one perceives but what one feels: atmospheres” [9]. As a consequence, conclusions drawn from the creative

experiments as well as experience gained from the installations and tools were used to develop elements of a didactic concept in which sensual perception of the phenomena of the interactions of color and light are linked to the direct experience of learners.

Selected Instruments of a Theory of Color and Light

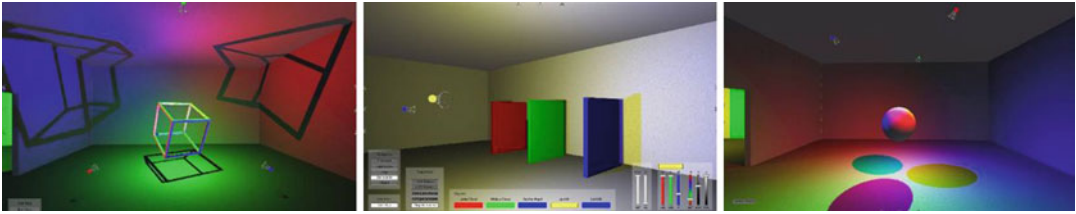
Installations

The authors identify as installations such arrangements as permit different perceptual phenomena relating to color and light to be placed in a spatial context so that they can be experienced by the senses. They evolve from observations such as can also be made in everyday life. In contrast, the phenomena thus presented are detached from the context of their surroundings and then staged



Interactions of Color and Light, Fig. 2 Color-light playground tool; project: *Colour and Light – Materials for a Theory of Colour and Light*. This tool stimulates experimentation with dynamically controlled LEDs and color-coated angled sheet metal which can be assembled

in a variety of combinations of color and space. Hue, saturation, and brightness can be altered manually, enabling active experience of the interactions between surface colors and colored light



Interactions of Color and Light, Fig. 3 The color-light toy

and varied spatially, such that these aspects, or some of them, can be perceived in isolation from it. In this process *hue*, *saturation*, and *brightness* of the colors observed are affected to varying degrees (Fig. 1).

Tools

The term “tools” refers to materials which facilitate play-based and experimental access to various perceptual phenomena around light and color. These include physical as well as digital tools. They encourage independent or directed experimentation and thus facilitate the acquisition of experiential knowledge in individualized learning processes (Fig. 2).

Color-Light Toy

The color-light toy is a virtual 3D environment where complex interactions of surface colors and light and various contrast effects can be simulated interactively (Fig. 3).

Conclusion

The authors’ research demonstrates the lack of integrative principles and practices in the field of designing using color and light, especially in the context of architecture, *interior architecture*, design, and *scenography*. The research demonstrates the high degree of willingness to apply the phenomena of interaction of color and light professionally and as end users. The approach of implementing the abovementioned phenomena into a mediation concept which can be experienced and experimented on, and which is focused on aspects of direct and sensual perception, is

being further developed by researchers into color and light at various institutions [10].

Cross-References

- [Architectural Lighting](#)
- [Color Combination](#)
- [Complementary Colors](#)
- [Color Contrast](#)
- [Environmental Color Design](#)
- [Inorganic, Hybrid, and Functional Pigments](#)
- [Light-Emitting Diode, LED](#)
- [Primary Colors](#)

References

1. Colourlight-Center: Zurich University of the Arts. www.colourlight-center.ch (2013)
2. Research Projects at the Zurich University of the Arts: ColourLightLab, supported by SNF/DORE, Swiss National Science Foundation (2005–2006); LED-ColourLab supported by KTI, Swiss Committee for Technology and Innovation (2008–2009); Colour and light – materials for a theory of colour and light, supported by SNF/DORE, Swiss National Science Foundation (2010–2011). www.colourandlight.ch and www.colourlight-center.ch
3. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications/International Educational and Professional Publisher, Thousand Oaks/London/New Delhi (2008)
4. Jonas, W.: Forschung Durch Design. In: Erstes Design Forschungssymposium. Swiss Design Network (2004)
5. Jonas, W.: Design research and its meaning to the methodological development. In: Michel, R. (ed.) Design Research Now. Birkhäuser Verlag AG, Basel/Boston/Berlin (2007)
6. Bachmann, U.: Colours Between Light and Darkness. Niggli Publisher, Sulgen (2006)

7. Bachmann, U.: Colour and Light – Materials for a Theory of Colour and Light. Niggli Publisher, Sulgen (2011)
8. Exposition, LED – Licht und Farbe inszenieren, Gewerbemuseum Winterthur (2008–2009)
9. Böhme, G.: Atmosphäre, p. 15. Suhrkamp, Frankfurt (1995)
10. Institute of Colour and Light, Zurich, www.colourandlight.ch; Institute Integrative Design, Academy of Art and Design Basel, www.masterstudiodesign.ch; Colourlight-Center, Zurich University of the Arts, www.colourlight-center.ch

Definition

Lighting concepts for interior spaces can be found everywhere, be it the concept to allow daylight to penetrate in a building without causing glare or overheating, be it the concept to create appropriate lighting conditions in offices, or be it the specifics of museum lighting to highlight art without damaging it by radiation or heat. Interior lighting is about choosing the appropriate lighting components in a lighting design to fit the requirements of a certain application (examples in Fig. 1).

Interior Lighting

Martine Knoop
Berlin University of Technology, Berlin,
Germany

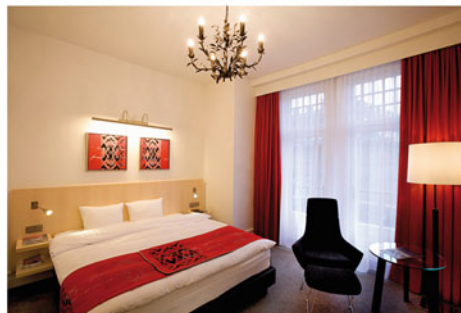
Synonyms

Indoor lighting; Lighting for indoor spaces; Lighting for interior spaces

Overview

Interior lighting can be applied to:

- Support safety from a visual point of view
- Provide adequate lighting conditions for activities in terms of quantity and quality
- Enhance the desired room appearance/lit environment
- Create an atmosphere suitable for the space that is lit



Interior Lighting, Fig. 1 Examples of interior lighting. (All pictures: Philips Lighting)

- Enhance health and well-being
- Support an architectural design

The first five aspects are related to the individuals (individual well-being) using the space; the last aspect is related to the space itself. Lighting for safety and lighting for performing activities need to be considered as minimum lighting conditions. This so-called functional lighting is typically referred to in standards, whereas aesthetic lighting, clearly related to atmosphere creation and support of architectural designs, is less specified. Depending on the application, these aspects will vary in their emphasis. For example, atmosphere creation is of higher importance in restaurants and shops; enhancing well-being with lighting is typically taken into consideration in lighting designs for, e.g., schools, hospitals, or office buildings.

Safety needs in normal operation, for example, to safeguard not to stumble over obstacles in a corridor or to safely operate industrial machines, are usually fulfilled when lighting conditions to perform activities in the space are realized. Safety in temporary, untypical situations, mainly in case of electrical power loss, needs to be addressed as well. In this case, the lighting needs to guarantee that the building can be vacated safely without electrical energy in a given time frame. Requirements for the so-called emergency lighting can be found in, for example, EN 1838:2013 Lighting applications – Emergency lighting and CIE S 020/E:2007 Emergency Lighting [1, 2].

Lighting to realize adequate conditions for activities focuses on the lighting needs for visual performance and visual comfort. Aspects to be considered are, for example, protection from glare caused by either daylight or electric lighting in office rooms, minimum lighting levels to perform visual tasks in almost all applications, color rendering requirements to enable differentiation between colors, which is specifically important, for instance, for color inspection, and in a large number of areas in health-care premises.

Lighting can be applied to enhance room appearance. Several studies indicate that people prefer relatively bright vertical surfaces [3–5].

Research [6] has indicated that the appearance of a space is related to the lighting characteristics of a 40° band of view for an observer (20° above and 20° below the line of sight). The luminance of this band has a large effect on the perception of dim-bright and uniform-nonuniform. A room is perceived bright from an average luminance of 30–40 cd/m² [6, 7, 8]. Additionally, the average scene luminance and ceiling luminance are also relevant parameters in the satisfaction rating of a lighting system [9].

Lighting to create an atmosphere sets the right scene for the purpose of the interior space. In retail and hospitality applications, lighting is an influential parameter to create an inviting space, to stimulate communication or to attract shoppers. Research has been conducted in the 1970s [10] and more recently [6, 11, 12, 13]. It indicates, for example, that to realize an intimate atmosphere, diffuse warm lighting should be applied; a businesslike, neutral atmosphere is created with diffuse cold lighting. Visual clarity is achieved with high workplane and wall luminances. A place works confined, narrowing when direct overhead lighting is applied at low lighting, without illuminating the walls. Nonuniform, moderate light levels and nonuniform wall lighting induce pleasantness.

Light is important to people's health and well-being. It can affect people's physiological and psychological state. Although already researched in the 1980s and 1990s, it received a broader interest after the discovery of a new photoreceptor (ipRGC) in the retina of the eye [14]. When light reaches these cells, signals are sent to the central biological clock (suprachiasmatic nuclei – SCN), which regulates circadian and circannual rhythms of a large variety of bodily processes. This influences, for example, sleepiness and alertness. The lighting community is still exploring how light can be used to support the healing process in hospitals, to improve learning conditions for pupils in schools, or to improve shift workers' sleep quality during the day and alertness during the night. Current research indicates that higher illuminances and a wider range of color temperatures than typically specified in standards can stimulate people and enhance their well-being

(► [Non-visual Lighting Effects and Their Impact on Health and Well-Being](#)).

Lighting can also be used to support the architectural design, by using coves and large lit surfaces to define the space, as well as by highlighting architectural details, such as columns, arches, and galleries (► [Architectural Lighting](#)).

Lighting Quality

Although no precise definition for lighting quality exists [15], lighting quality is typically related to the degree of which the lighting conditions meet needs of users of the space. A large field investigation in office buildings [9] has indicated that the effect of lighting that is perceived of higher quality can be considerable. The study shows that, through environmental satisfaction, lighting quality can affect job satisfaction, organizational commitment, and turnover, as well as health problems and self-assessed performance.

In order to realize good quality lighting, the lighting solution needs to fulfill the requirements set by/coming from the various target groups, for example, the user, the owner, and/or the designer. The quality of the lighting solution is determined by the fulfillment of the set lighting purpose(s) (see before) as well as the economic, environmental, and design aspects of the solution, such as style, integration into architecture, appropriateness of appearance, form, and size (from [5, 16]). Quality lighting can be achieved by the right choice of lighting components (light sources, luminaires, and lighting control) in an appropriate lighting design.

This part on interior lighting will focus on quality aspects of lighting conditions for activities, as mentioned above. Typically, requirements for this lighting can be found in standards and recommendations (e.g., [15, 17–20]), which ensure that the user in the application is able to comfortably perform basic visual work. The following aspects are usually seen as minimum considerations.

Task Visibility

In order to have appropriate task visibility in indoor spaces, a minimum required (average) illuminance level is given in standards and recommendations. Just as an example, in the European standard EN 12464-1 [18], a minimum average illuminance of 500 lx is given for writing, typing, reading, and data processing in office buildings; color inspection in chemical, plastics, and rubber industry requires 1,000 lx. A minimum uniformity of illumination ($E_{\min}/E_{\text{av}}$) is demanded to prevent frequent reaccommodation of the eyes. In the European Standard the required uniformity lays between 0.4 (for an easy task or in the surrounding area of a task) and 0.7 (difficult task).

In order to prevent errors, fatigue, or accidents, one should ensure that reflections in specular surfaces (veiling reflections) are not present. These reflections, which partially or totally obscure the details to be seen by reducing the contrast, will diminish task visibility. Maximum average luminance values are given to avoid reflected glare on computer screens [18]. When it comes to reflections in glossy material (e.g., magazine) or glass (e.g., vitrine or pictures) (see Fig. 2), design rules that define the appropriate position of the light sources (outside the so-called offending zone)



Interior Lighting, Fig. 2 Examples of veiling reflections

can be applied to prevent veiling reflections (see, e.g., [21]).

Visual Comfort

When too high luminance differences are offered within the field of view, glare can be caused. Discomfort glare is the type of glare that does not directly reduce the ability to see an object, but it produces a sensation of discomfort and might affect performance in the longer run. Glare which impairs vision is called disability glare. In indoor lighting it is assumed that disability glare does not occur, when both shielding angles for bright light sources and discomfort glare restrictions are applied. The degree of discomfort glare can be assessed with the Unified Glare Rating or Visual Comfort Probability. Standards and recommendations provide maximum (UGR) and minimum (VCP) values to ensure comfortable lighting conditions in a specific application. For example, for offices the required UGR is typically 19, the recommended VCP value is 70. For situations with VDT tasks, a VCP value of 80 is recommended (► [Glare](#)).

Whereas the task visibility and visual comfort aspects are related to the lighting solution in a specific application, standards and recommendations refer to minimum requirements for characteristics of light sources as well. Both flicker and color rendering are qualities to be assessed, as they might hamper visual performance when requirements are not met.

Visible, but also invisible, flicker can trigger headaches, fatigue, and distraction. It has been shown to reduce performance. If it results in stroboscopic effects and objects appear to move discretely rather than continuously, it can lead to critical situations, e.g., in industry when interacting with rotating or reciprocating machinery. The Critical Flicker Frequency and the Flicker Index are used in the past to assess flicker, but research indicates that not all relevant parameters are covered in both measures [22–24]. New metrics are looked into, to be used to predict both flicker and stroboscopic effects for all light sources, including LED (e.g., [25–28]).

For visual performance, colors in the environment shall be rendered naturally and correctly.

The general color rendering index (R_a) of a light source is used as an indication of how well this light source reproduces colors of an object in comparison with a reference (ideal or natural) light source. The maximum value of R_a is 100, and the required value is depending on the application. For example, according to the European Standard EN 12464-1 [18], indoor circulation areas require a minimum R_a of 40, offices a value above 80, and areas for fabric control in textile manufacturing or examination and treatment rooms a minimum R_a of 90. The color rendering index of a light source can be obtained from the manufacturer, the package, or the lamp itself (► [CIE Color-Rendering Index](#)). The R_a has limitations, especially for LED light sources, and other metrics to describe color quality are developed for practical application (e.g., in [29]) and usage in research [30].

A number of quality measures go beyond task performance. Generally, these are listed in standards or recommendations as well, with proposed minimum levels to achieve a higher lighting quality, to benefit human health and well-being (e.g., in [18, 20]). Examples are:

- Color appearance, which is referring to the color of the light, quantified by its correlated color temperature (T_{CP}). For electric lighting, it ranges typically from 2,700 K (warm white light) to 6,500 K (cool white light), with a maximum of 17,000 K. The choice for a certain color appearance depends on the application, the atmosphere to be created, the colors of furniture and room partitions, the illuminance level, climate, and culture.
- Modelling of faces for visual communication and recognition of faces. In order to achieve this, the lighting in a room should not be too directional nor too diffuse. Design guidelines propose to use as little as 20% of the total illuminance from a diffuse or reflected light source and avoid strongly directional lighting (e.g., [19, 20, 31]).

It needs to be noted that the abovementioned quality criteria typically focus on minimum requirements to assure visual performance and

visual comfort. As indicated before, room appearance is of great importance for the user's well-being. The standards, mainly focusing on minimum requirements for task illumination, will not sufficiently address this. Although minimum required vertical illuminance is mentioned in the IES Handbook [16] and the EN 124641:2011 [18], the needed lighting levels for quality lighting addressing also room appearance and people's health and well-being are likely higher.

Interior Lighting Design

Interior lighting is about choosing the lighting components in a lighting design to fit the requirements of a certain application.

As indicated before, a lighting design consists of functional lighting and/or aesthetic lighting. Typical application areas with a focus on functional lighting are industry (e.g., assembly work, storage) and office (e.g., open plan office, cell office, meeting room, corridors). Typical application areas with additional aesthetic lighting are retail (e.g., department stores, boutiques), hotels (e.g., entrance areas, reception areas), and restaurants.

Resulting, interior lighting is often built up from layers of light. Depending on the application, one or more layers are applied, typically distinguishing between

- **General lighting:** lighting providing (horizontal) illumination over the total area of the room with a certain degree of uniformity. General lighting is often offered in a regular grid of point, line, or area light sources (see Fig. 3) (► [“Coefficient of Utilization, Lumen Method,”](#) ► [“Diffuse Lighting”](#)).
- **Task lighting:** additional lighting to general lighting in the area in which the task will be performed. Task lighting is used to fulfill the requirements, without having to illuminate the entire room to that specific level, in order to reduce energy consumption for electric lighting (see Fig. 4) (Task Lighting).
- **Accent lighting:** directional lighting to emphasize a particular object or to draw attention to a part of the field of view (see Fig. 5).
- **Decorative lighting:** the lighting itself provides a point of interest or attraction in an interior (see Fig. 6).
- **Architectural lighting:** lighting to form or underline the architecture of the space (see Fig. 7) (► [Architectural Lighting](#)).

All layers have specific requirements, which the lighting components have to fulfill.

Lighting Components for Interior Lighting

Lighting for indoor places can be provided by daylight, electric lighting, or a combination of both. In this entry, only electric lighting is discussed. As interior lighting is about choosing the lighting components in a lighting design to fit the requirements of a certain application, the lighting components will be briefly discussed below.

Interior lighting is used in a wide spread of applications with a wide range of requirements. Resulting, a large variety of light sources and luminaire types are applied in interior lighting. The most relevant characteristics that determine the applicability of a light source in a certain application are luminous flux, efficacy, luminance, color rendering, correlated color temperature, lifetime, run-up time, reignition time, and form factor of the light source. In the design process the application requirements on these aspects need to be specified, in order to select the appropriate light source.

To give an impression, without being conclusive, one can roughly say that halogen lamps are generally applied in aesthetic lighting, in accent lighting, and in decorative lighting (► [Incandescent Lamp](#)). Fluorescent lamps can be found in applications with a focus on functional lighting, in both general lighting and task lighting, such as industry, health care, educational buildings, and office. Next to that fluorescent lamps are used in architectural lighting solutions,



Interior Lighting, Fig. 3 Examples for general lighting: point (two pictures on the top), line, and area light sources. (First picture: Philips Lighting)

such as coves and luminous ceilings (► [Tubular and Compact Fluorescent Lamp](#)). High-intensity discharge lamps (e.g., ► [Metal Halide Lamp](#)) can be found in general lighting solutions as well as accent lighting solutions, in, for example, industry and retail applications. LEDs have been successfully introduced in interior lighting as well and can be found in both functional and aesthetic lighting solutions. The diversity of LEDs allows for usage of this light source in general, task, accent, decorative, as well as architectural lighting (► [Light-Emitting Diode, LED](#)).

With respect to the luminaires (► [Luminaires](#)), a similar specification of requirement needs to be made. Based on the chosen required lighting concept (functional and aesthetic, from general lighting to architectural lighting), the following

characteristics of the luminaire can support luminaire selection: the light distribution (luminous intensity distribution – ► [“Light Distribution”](#)), efficiency (reflected in light output ratio), IP class, form factor, and type (recessed, surface mounted, suspended, wall mounted, free floor standing, or desk luminaire). The variety is large, serving the broad bandwidth of requirements in interior lighting.

In the overall lighting design, lighting controls need to be considered. Reasons to use lighting controls in interior lighting are to:

- Make use of potential energy savings, such as time control, occupancy control, and controls that respond to the daylight availability in the room (daylight supply control)



Interior Lighting, Fig. 4 Examples for task lighting. (All pictures: Philips Lighting)

- Increase user comfort and well-being, such as personal control, glare control, and controls that offer lighting settings for specific tasks (task demand control)
- Support building appearance, ambience, and company representation, for example, using scene setting

Lighting controls can enhance the quality of a lighting installation. Once more, the specification will depend on the application. The use of occupancy control is, for example, suitable in areas where people generally stay a relatively short period or do not feel responsible for switching

off the lighting, such as bathrooms, stair cases, and warehouse aisle. Daylight supply control is applied in situations with sufficient daylight and an appropriate choice, for example, in open plan offices and hospital wards (► [Lighting Controls](#)).

The combination of components – light source, luminaire, and controls – will determine the overall quality of the lighting.

Concluding

Interior lighting serves a number of objectives. Each application requires a specific approach,



Interior Lighting, Fig. 5 Example for accent lighting. (Picture: Philips Lighting)



Interior Lighting, Fig. 6 Examples for decorative lighting. (Second picture, Philips Lighting; photo credits, Cinimod Studio – third picture, Philips Lighting; photographer, Leon Verlaek)

typically build up from a few layers of light, fulfilling demands of the user of the space and its architecture. Good quality lighting respects

these demands, as well as the economic, environmental, and design aspects. A combination of suitable light sources, luminaires, and lighting



Interior Lighting, Fig. 7 Examples for architectural lighting. (Second and third picture: Philips Lighting)

controls in an appropriate design will offer good quality lighting in interior spaces.

Cross-References

- Architectural Lighting
- CIE Color Rendering Index
- Coefficient of Utilization, Lumen Method
- Diffuse Lighting
- Glare
- Incandescent Lamp
- Light Distribution
- Light-Emitting Diode, LED
- Lighting Controls
- Luminaires
- Metal Halide Lamp
- Nonvisual Lighting Effects and Their Impact on Health and Well-Being
- Tubular and Compact Fluorescent Lamp

References

1. CEN: EN 1838:2013 Lighting Applications – Emergency Lighting. Published by the European National Standards Institutes CEN, Brussels, Belgium (2013)
2. CIE: ISO 30061:2007(E)/CIE S 020/E:2007 Emergency Lighting. CIE, Vienna (2007)
3. Newsham, G.R., Marchand, R.G., Veitch, J.A.: Preferred surface luminances in offices, by evolution. *J. Illum. Eng. Soc.* **33**(1), 14–29 (2004)
4. Van Ooyen, M.H.F., Van De Weijgert, J.A.C., Begemann, S.H.A.: Preferred luminances in offices. *J. Illum. Eng. Soc.* **16**(2), 152–156 (1987)
5. Veitch, J.A.: Lighting for high-quality workplaces (NRCC-47631 – <https://nrc-publications.canada.ca/eng/view/accepted/?id=7ff04ffb-4adf-4db8-93c9-81eb297c6d9d>. Accessed 25 June 2019), also published in Clements-Croome, D. (ed.): *Creating the Productive Workplace*, 2nd edn, pp. 206–222. Taylor & Francis, London (2006)
6. Loe, D.L., Mansfield, K.P., Rowlands, E.: Appearance of lit environment and its relevance in lighting design: experimental study. *Light. Res. Technol.* **26**(3), 119–133 (1994)

7. Newsham, G.R., Arsenault, C., Veitch, J.A., Tosco, A.M., Duval, C.: Task lighting effects on office worker satisfaction and performance, and energy efficiency. *Leukos*. **F1**(4), 7–26 (2005)
8. Kirsch, R.M.: Lighting quality and energy efficiency in office spaces. Thesis, Technische Universität Berlin. <https://depositonce.tu-berlin.de/handle/11303/4735> (2015). Accessed 25 June 2019
9. Veitch, J.A., Newsham, G.R., Mancini, S., Arsenault, C.D.: Lighting and office renovation effects on employee and organizational well-being NRC report IRC – RR-306. <https://nrc-publications.canada.ca/eng/view/fulltext/?id=24aa7d6c-12b4-4fec-aa79-920d575d319c> (2010). Accessed 25 June 2019
10. Flynn, J., Spencer, T.J., Martyniuk, O., Hendrick, C.: Interim study of procedures for investigating the effect of light on impression and behavior. *J. Illum. Eng. Soc.* **3**, 94–96 (1973)
11. Vogels, I.: Probing experience atmosphere metrics: a tool to quantify perceived atmosphere. *Philips Res. Book Ser.* **8**(1), 25–41 (2008)
12. Custers, P.J.M., De Kort, Y.A.W., IJsselstein, W.A., De Kruijff, M.E.: Lighting in retail environments: atmosphere perception in the real world. *Light. Res. Technol.* **42**(3), 331–343 (2010)
13. Wang, H.H., Luo, M.R., Liu, P., Yang, Y., Zheng, Z., Liu, X.: A study of atmosphere perception of dynamic coloured light. *Light. Res. Technol.* **46**(6), 661–675 (2014)
14. Berson, D.M., Dunn, F.A., Takao, M.: Photo-transduction by retinal ganglion cells that set the circadian clock. *Science*. **295**(5557), 1070–1073 (2002)
15. Veitch, J.A. (ed.): CIE: Proceedings of the First CIE Symposium on Lighting Quality. CIE Central Bureau, Vienna (1998)
16. DiLaura, D., Houser, K., Mistrick, R., Steffy, G. (eds.): IESNA: Lighting Handbook: Reference & Application, 10th edn. IESNA, New York (2011)
17. CIE: CIE S 008/ISO 8995-1: Lighting of Work Places – Part 1: Indoor. CIE, Vienna (2002)
18. CEN: EN 12464-1:2011 Light and Lighting – Lighting of Work Places – Part 1: Indoor Work Places. Published by the European National Standards Institutes CEN. Brussels, Belgium (2011)
19. IESNA: ANSI/IESNA RP-1-04 American National Standard Practice for Office Lighting. IESNA, New York (2012)
20. IESNA: IES DG-18-08 A Guide to Designing Quality for People and Buildings. IESNA, New York (2009)
21. Ganslandt, R., Hofmann, H.: Handbook of Lighting Design. ERCO Edition, Verlag Vieweg ISBN 3-528-08895-8. <https://www.erco.com/download/content/4-media/2-handbook/erco-handbook-of-lighting-design-en.pdf> (1992). Accessed 25 June 2019
22. Vogels, I., Sekulovski, D., Perz, M.: Visible artefacts of LEDs conference proceedings of the 27th session of the CIE. *Sun City*. **1**(1), 42–51 (2011)
23. Bullough, J.D., Sweater Hickcox, K., Klein, T.R., Narendran, N.: Effects of flicker characteristics from solid-state lighting on detection, acceptability and comfort. *Light. Res. Technol.* **43**(3), 337–348 (2011)
24. Bullough, J.D., Hickcox, K.S., Klein, T.R., Lok, A., Narendran, N.: Detection and acceptability of stroboscopic effects from flicker. *Light. Res. Technol.* **44**(4), 477–483 (2012)
25. ASSIST Alliance for Solid-State Illumination Systems and Technologies: ASSIST Recommends ... Flicker Parameters for Reducing Stroboscopic Effects from Solid-State Lighting Systems, vol. 11, no. 1. Lighting Research Center, Troy. <http://www.lrc.rpi.edu/programs/solidstate/assist/pdf/AR-Flicker.pdf> (2012). Accessed 25 June 2019
26. Bodington, D., Bierman, A., Narendran, N.: A flicker perception metric. *Light. Res. Technol.* **48**(5), 624–641 (2016)
27. Perz, M., Sekulovski, D., Vogels, I., Heynderickx, I.: Quantifying the visibility of periodic flicker. *Leukos*. **13**(3), 127–142 (2017)
28. Perz, M., Sekulovski, D., Vogels, I.M., Heynderickx, I.E.: 77-1: Invited paper: modelling visibility of temporal light artefacts. In: *SID Symposium Digest of Technical Papers*, vol. 49, no. 1, pp. 1028–1031. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/sdtp.12194> (2018). Accessed 25 June 2019
29. IES: IES TM-30-15 IES Method for Evaluating Light Source Color Rendition. IES, New York (2015)
30. CIE: CIE 224:2017 Colour Fidelity Index for Accurate Scientific Use. CIE, Vienna (2017)
31. Boyce, P.R., Miller, N.J., Wilwol, K.: Lighting for good facial modelling. <https://www.lrc.rpi.edu/resources/pdf/68-1999.pdf> (1999). Accessed 25 June 2019

Internet of Things and LEDs

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

IoT

Definition

IoT is a network that connects uniquely identifiable “things” to the Internet. The “things” have sensing/actuation and potential programmability capabilities. Through the exploitation of unique

identification and sensing, information about the “thing” can be collected, and the state of the “thing” can be changed from anywhere, anytime, by anything [1]. LED lighting systems can be used both as the network for IoT systems and as the carrier for the required sensors and activators.

Internet of Things (IoT)

In more laymen terms, the definition of Internet of things can be described as a system that:

- connects us with physical devices (“things”) around us,
- collects and stores information (data) from these devices,
- smartly controls the devices around us based on the collected information.

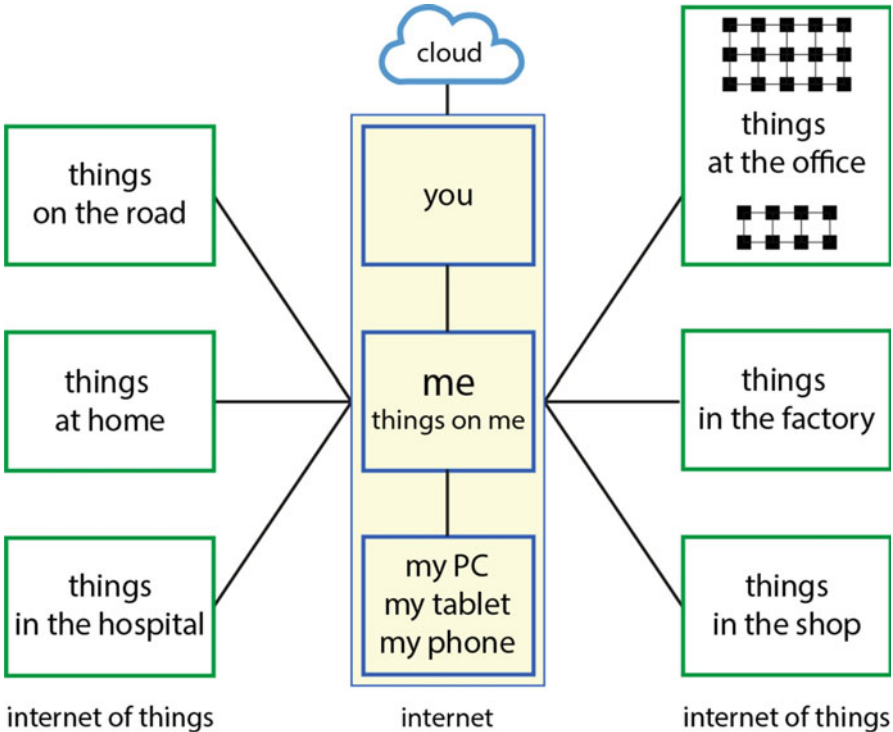
A typical example of an IoT system is the self-driving communicative car. With this system, my

car communicates with other cars, collects information from them (including steering and braking actions), analyzes the collected data, and decides how to react and executes the best maneuver.

IoT needs four fundamental components:

- Sensors (getting data)
- Activators (executing commandos)
- Network connection
- Data processing (analyzing)

Figure 1 shows the relationship between the Internet and the Internet of things (IoT). Internet connects me with other persons (you) and with our PC, tablet, and smartphone. In different environments, such as our home, offices, factories, shops, roads, and hospitals, many IoT-enabled devices which incorporate sensors and activators are interconnected in local networks, which, in turn, can be connected to the Internet. Some sensors may be located on ourselves. Think of smartwatches which sense heartbeat, activity,



Internet of Things and LEDs, Fig. 1 Internet and Internet of things (IoT) connections

and many other things. Perhaps sounding futuristic, some persons will have chipped sensors which directly connect with their medical professionals. As an example, for the office environment, two different networks, connected with the Internet, are drawn. These could, for instance, represent a lighting network with integrated sensors and activators and an HVAC network with integrated activators.

Microcontrollers, activators (e.g., for switching and dimming), and sensors, such as light, occupancy, temperature, humidity, and noise sensors, are small enough to be incorporated in the LED luminaire from which they also get the power needed for their operation. The sensors used in the smart lighting network can be expanded with sensors that collect information, such as occupancy, temperature, humidity, and air quality needed to control and optimize other installations in the building. The smart lighting network so becomes the “heart” of the Internet of things of that building. An IoT lighting system can considerably increase the energy efficiency of the electric installations in a building and reduce the costs of operating that building while ensuring optimized performance, comfort, and health of the building users.

IoT Lighting System

Infrastructure for IoT Lighting Systems

Crucial in IoT systems are the sensors that collect information from the things around us. IoT systems need a granular sensory network permitting sensors to be ubiquitous (omnipresent) and densely spaced and give them access to power. Such a network can “sense what is going on” at each spot of the space. Fortunately, there is no need for a new infrastructure for such a sensory network, as the lighting network fulfils the above requirements. The network of light points can serve as the infrastructure for IoT systems. This is not only the case for indoor lighting in the context of smart buildings but also for road lighting in the context of smart cities (Fig. 2). By having the sensors, or most of the sensors, integrated with the luminaires

themselves, the lighting system becomes an IoT lighting system.

IoT Lighting System Technology

Data received from the sensors can be stored and processed in the cloud. Local storing and processing of some of the data can be advantageous to get actions, as a result of sensed changes, fast enough for security reasons and to avoid failures of the system in case of bad network connections. The local storing and processing at the borders (edges) of the own network is in the IoT world referred to as “edge computing.” Usually, a combination of cloud and edge storing and computing is used in modern IoT systems. The quality of the analysis of the large amount of data received determines the effectiveness of the resulting actions of the IoT system. Machine learning, a field of artificial intelligence (AI) that uses complex statistics with which computer systems “learn” to improve their effectiveness continuously, offers important possibilities in this respect.

Incompatibility of connected devices of different manufacturers and the use of different, propriety or non-propriety, communication standards is a problem for interoperability. The European Union has therefore initiated the project OpenAIS (Open Architectures for Intelligent Solid-State Lighting Systems). It has defined an interoperability standard allowing different systems to work collaboratively [2, 3]. The project has been carried out by the OpenAIS consortium, existing of representatives of the European lighting industry and two academic institutions [4]. The consortium develops components and standards to connect all luminaires to the Internet. Technological advances and changes in working habits and regulations all play a role in the work of the consortium.

Services of IoT Lighting Systems

The data collected by the IoT system are available in real time as input for direct, smart actions of the connected building service installations such as lighting, heating, ventilation, and air conditioning (HVAC). Since the data are stored, they are also



Internet of Things and LEDs, Fig. 2 The network of light points, in buildings and on roads, is a granular network that is ubiquitous and densely spaced and has access to power and can, therefore, serve as the infrastructure for IoT systems

available for data-driven decisions by facility management such as space allocation, cleaning, and maintenance actions and for helping with diagnostics. Part of these decisions can be automated. For general management, the data can be used for the optimization of energy consumption and costs.

By following in real time the occupation of spaces suitable for different activities, new areas

can be made automatically available when needed. The not-needed spaces need no lighting and less heating and air conditioning. By having the system analyze the historical occupation data, individual rooms may be closed, again automatically, certain days of the week with all services completely switched off. The cleaning and maintenance schedule can be adapted to the actual occupancy situation.

Security management can be facilitated by automatic identification of persons with the aid of smart beacons installed at suitable places, including the entrances. These beacons communicate through Bluetooth with an app on the smartphones of the persons having access to a building. The same system can be used to locate a person or persons in an emergency and send short messages. Again the same method can be used to “tell” the luminaires and HVAC system which person is in the influencing area of them. In this way, the person, wherever he is in the building, can control the services of light, heating, or air conditioning around him with his personal control device, for example, his smartphone. Or, even more advanced, the services automatically sense the personal preferences of the person and switch the services accordingly. If desired, the system can be made to include the coffee or tea machine serving automatically coffee or tea according to the personal taste.

Customer relation management (CRM) by shops owners can benefit from IoT systems as well. By installing beacons in the shop and providing an app for customers, shoppers can be monitored as far as locations visited are concerned. This information can be compared with the CRM base, which, of course, includes what products are bought previously by the customer. Special offers can be custom-made, right on the spot. The next section discusses that such a shop app can also use the smart luminaires as the basis for an indoor navigation system so that the customer can easily find the products he is looking for.

Many IoT lighting systems realized in smart commercial buildings, using different kinds of data, are described in the literature.

IoT systems using the road lighting network as the carrier may become the backbone of smart cities. Sensors, built into the road lighting luminaires, can be used to control the road lighting itself, for example, in dependence of traffic speed and density, daylight availability, rain, fog, and frozen roads. Since information on the status of the LED light sources itself can also be made available, road lighting maintenance can be optimized. And, the same system can also be used to optimize the mobility in cities (and on motorways) by smart traffic and parking guidance

based on the information received from the sensors at different locations. Finally, the quality of life can be monitored and controlled to some extent. Think here of security, crowd control, and air quality control.

Under the entry “Visible Light Communication,” it is shown that LED light itself can be used as a sensor: LED lighting beyond illumination alone.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Visible Light Communication](#)

Acknowledgment Adapted by permission from Springer Nature “Interior Lighting, fundamentals, technology and application” by Wout van Bommel (2019)

References

1. Minerva R, Biru A, Rotondi D (2015) Towards a definition of the Internet of Things (IoT), Revision 1. Published by IEEE
2. Mathews, E., Guclu, S.S., Liu, Q., Ozcelebi, T., Lukkien, J.J.: The Internet of Lights: An Open Reference Architecture and Implementation for Intelligent Solid State Lighting Systems. *Energies*. **10**(1187), 1–27 (2017)
3. European Commission (2018) Open AIS D7.6 Final report of project Architectures for Intelligent solid state lighting systems
4. OpenAiS(2018) Open Architectures for Intelligent Solid State Lighting Systems <http://openais.eu/en/consortium>. Accessed on 17 October 2018

Interpolation of Spectral Data

Stephen Westland
Colour Science and Technology, University of
Leeds, Leeds, UK

Synonyms

[Subsampling](#)

Definition

Spectral data can be interpolated using Sprague interpolation when the original data are evenly spaced. This involves fitting a fifth-order polynomial between two points P_i and P_{i+1} thus:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 \quad (1)$$

which can be used to interpolate between the points P_i and P_{i+1} and where x is the proportional distance between P_i and P_{i+1} of the point to be evaluated. However, the polynomial is constrained so that the gradients and curvature are consistent with the points P_{i-2} , P_{i-1} and P_{i+2} , P_{i+3} (the two points either side of the points that are being interpolated between). This constraint results in the coefficients of equation being easily calculated from the points P_{i-2} – P_{i+3} [1] thus:

$$\begin{aligned} a_0 &= P_i \\ a_1 &= (2P_{i-2} - 16P_{i-1} + 16P_{i+1} - 2P_{i+2})/24 \\ a_2 &= (-P_{i-2} - 16P_{i-1} - 30P_i + 16P_{i+1} - P_{i+2})/24 \\ a_3 &= (-9P_{i-2} + 39P_{i-1} + 70P_i + 66P_{i+1} \\ &\quad - 33P_{i+2} + 7P_{i+3})/24 \\ a_4 &= (13P_{i-2} - 64P_{i-1} + 126P_i - 124P_{i+1} \\ &\quad + 61P_{i+2} - 12P_{i+3})/24 \\ a_5 &= (-5P_{i-2} + 25P_{i-1} - 50P_i + 50P_{i+1} \\ &\quad - 25P_{i+2} + 5P_{i+3})/24 \end{aligned} \quad (2)$$

However, for N points P_i where $i = 1:N$ the method cannot be directly used to interpolate between P_1 and P_2 nor between P_{N-1} and P_N because the additional points on either side do not exist. The CIE recommends a method for extrapolating four additional points, two at shorter wavelength intervals than available (P_{-1} and P_0) and two at longer wavelength intervals than available (P_{N+1} and P_{N+2}) thus:

$$\begin{aligned} P_0 &= (884P_1 - 1960P_2 + 3033P_3 - 2648P_4 \\ &\quad + 1080P_5 - 180P_6)/209 \\ P_1 &= (508P_1 - 540P_2 + 488P_3 - 367P_4 + 144P_5 \\ &\quad - 24P_6)/209 \\ P_{N+1} &= (-24P_{N-5} + 144P_{N-4} - 367P_{N-3} \\ &\quad + 488P_{N-2} - 540P_{N-1} + 508P_N)/209 \\ P_{N+2} &= (-180P_{N-5} + 1080P_{N-4} - 2648P_{N-3} \\ &\quad + 3033P_{N-2} - 1960P_{N-1} + 884P_N)/209 \end{aligned} \quad (3)$$

Background

CIE tristimulus values can be calculated from spectral reflectance factors, spectral illuminant relative power distributions, and the CIE color-matching functions (CMFs). However, it is sometimes the case that the CMFs, for example, are available at intervals of 5 nm, but the spectral reflectance factors are available at intervals of 10 nm. Interpolation of spectral data may be required and interpolation methods therefore need to be considered.

A straight line can be drawn to fit exactly through any two points and a parabola through any three points. More generally, an n th-degree polynomial can pass through any $n + 1$ points. Imagine we wish to fit some spectral data $P(\lambda)$ at four wavelengths, λ_1 – λ_4 . We can write:

$$\begin{aligned} P(\lambda_1) &= a_1\lambda_1^3 + a_2\lambda_1^2 + a_3\lambda_1 + a_4 \\ P(\lambda_2) &= a_1\lambda_2^3 + a_2\lambda_2^2 + a_3\lambda_2 + a_4 \\ P(\lambda_3) &= a_1\lambda_3^3 + a_2\lambda_3^2 + a_3\lambda_3 + a_4 \\ P(\lambda_4) &= a_1\lambda_4^3 + a_2\lambda_4^2 + a_3\lambda_4 + a_4 \end{aligned} \quad (4)$$

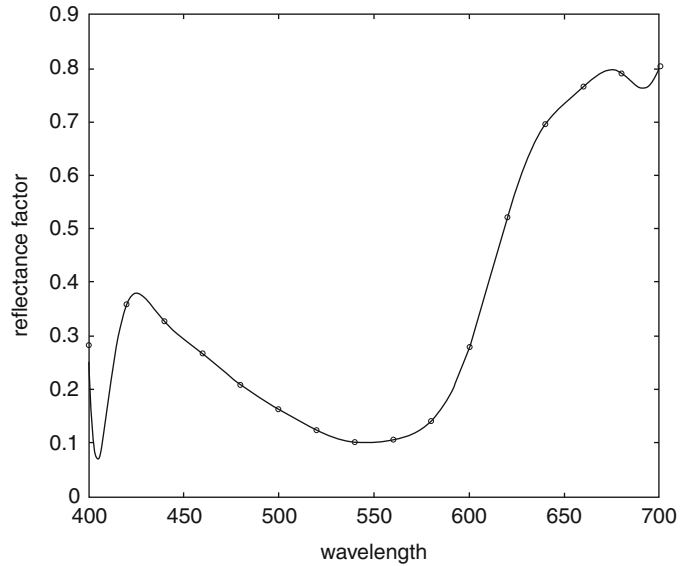
which represents four simultaneous equations and four unknowns. In terms of linear algebra, Eq. 4 can be represented by Eq. 5:

$$\mathbf{A}\mathbf{p} = \mathbf{q} \quad (5)$$

where $\mathbf{q}$ is a 4×1 column matrix of reflectance values, $\mathbf{p}$ is a 4×1 column matrix containing the coefficients a_1 – a_4 , and $\mathbf{A}$ is a special 4×4 matrix

Interpolation of Spectral Data,

Fig. 1 Example of interpolating 20-nm spectral data (symbols) using a Lagrangian polynomial (solid line)



known as the Vandermonde matrix. In the example shown as Eq. 4, the Vandermonde matrix would be constructed with the entries thus:

$$\begin{pmatrix} \lambda_1^3 & \lambda_1^2 & \lambda_1 & 1 \\ \lambda_2^3 & \lambda_2^2 & \lambda_2 & 1 \\ \lambda_3^3 & \lambda_3^2 & \lambda_3 & 1 \\ \lambda_4^3 & \lambda_4^2 & \lambda_4 & 1 \end{pmatrix}$$

The polynomial in Eq. 4 is referred to as the Lagrange polynomial. It is trivial to solve Eq. 5 for the coefficients in **p** using MATLAB thus **p** = **A**\b or, alternatively, using MATLAB's *polyfit* function. However, the use of a Lagrangian polynomial in this way is not the best way to interpolate spectral data; the resulting polynomial may lead to wide excursions between the points [1].

Figure 1 illustrates a Lagrangian polynomial that was fitted to spectral data at intervals of 20 nm and interpolated at intervals of 1 nm. The interpolation is likely to be inadequate at the short and long ends of the spectrum.

One way to address this problem is to use a family of polynomials (each of which fits a relatively small number of points) rather than trying to fit the whole spectrum with a single polynomial. Repeated application of a low-order polynomial

to a restricted range of the data is an example of a "piecewise method." CIE Publication Number 15.2 recommends that interpolation of spectra be carried out using a third-degree polynomial from neighboring data within twice the measurement interval [2]. This means that a Lagrange interpolation formula should be used for four data points (two either side of the point to be interpolated). A later CIE publication [3] recommends one of four interpolation methods: (1) the third-order polynomial interpolation (Lagrange) from the four neighboring data points around the point to be interpolated, (2) a cubic spline interpolation formula, (3) a fifth-order polynomial interpolation from the six neighboring data points around the point to be interpolated, or (4) a Sprague interpolation.

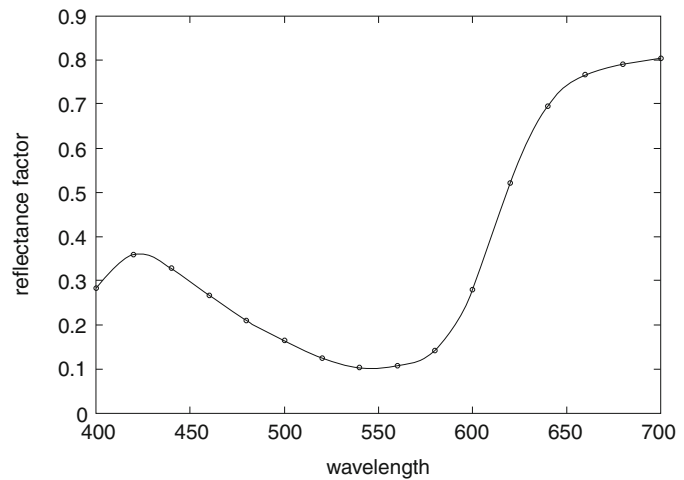
Current Status

Sprague interpolation was recommended following a study by a CIE working group [1], and the method for doing this is defined by Eqs. 1 and 2. Figure 2 illustrates some spectral data at intervals of 20-nm Sprague interpolation at intervals of 1 nm.

The method can be used equally for reflectance factors or for illuminant spectral power distributions.

Interpolation of Spectral Data,

Fig. 2 Example of interpolating 20-nm spectral data (symbols) using the Sprague method (solid line)



Cross-References

► [CIE Tristimulus Values](#)

References

1. CIE Publication No. 167: Recommended practice for tabulating spectral data for use in colour, Central Bureau of the Commission Internationale d'Eclairage (2005)
2. CIE Publication No. 15.2: Colorimetry, 2nd edn, Central Bureau of the Commission Internationale d'Eclairage (1986)
3. CIE Publication No. 15: Colorimetry, 3rd edn, Central Bureau of the Commission Internationale d'Eclairage (2004)

Interreflection

► [Global Illumination](#)

Intrinsic Decomposition

► [Intrinsic Imaging](#)

Intrinsic Images

► [Intrinsic Imaging](#)

Intrinsic Imaging

Elena Garces
Multimodal Simulation Lab (MSLab),
Universidad Rey Juan Carlos, Madrid, Spain

Synonyms

[Intrinsic decomposition](#); [Intrinsic images](#)

Definition

Intrinsic imaging or intrinsic image decomposition is the problem of decomposing an image into two layers: the reflectance layer, which describes the albedo color of the material; and the illumination or shading layer, which contains information about the interaction between light and geometry. This decomposition is physically accurate for Lambertian materials, which reflect light uniformly independent on the observer's position. Obtaining this decomposition for single images is a highly ill-posed under-constrained problem, as for each known value of the input image, there are two additional unknowns. While still an open problem, since the 70s many methods have tried to tackle this problem by making assumptions or imposing priors over the layers. These priors might come from

different sources: studies of the human perception, statistics of natural images, user interaction, or directly learned from the data.

Overview

The appearance of an object in the physical world is determined by the complex interaction between the light, the material, and the geometry of the scene that contains it. Capturing this information in isolation for post-processing operations such as relighting or re-texturing can be done with different setups. The illumination is most frequently obtained by taking a light probe, which is 360° HDR picture of the environment. The material properties can be captured with different devices. Among such devices, the *goniophotometer* is the most accurate one to obtain the radiance under all possible combinations of incoming and outgoing light and view directions, capturing the behavior of the material in the form of Bidirectional Texture Functions. Finally, geometry can be obtained accurately with stereo cameras or depth sensors. These setups, however, require the use of specific devices and planning, and are not useful for on-the-wild situations where the only available device at hand is a conventional photographic camera.

When a light photon hits a pixel of the camera sensor, it integrates all the information that conforms the scene –geometry, materials and illumination– in a single value, making the problem of recovering back this data an ill-posed one. A typical strategy to obtain this latent information involves fitting an analytical function parameterized by geometry, light and material, problem which is known as *inverse rendering*. This estimated model will allow to fully modify and *re-render* the scene in post-processing, e.g., changing the illumination or viewpoint. The complexity of such model will determine the degree of realism of the matched scene, along with the difficulty in estimating its parameters, and the editing capabilities.

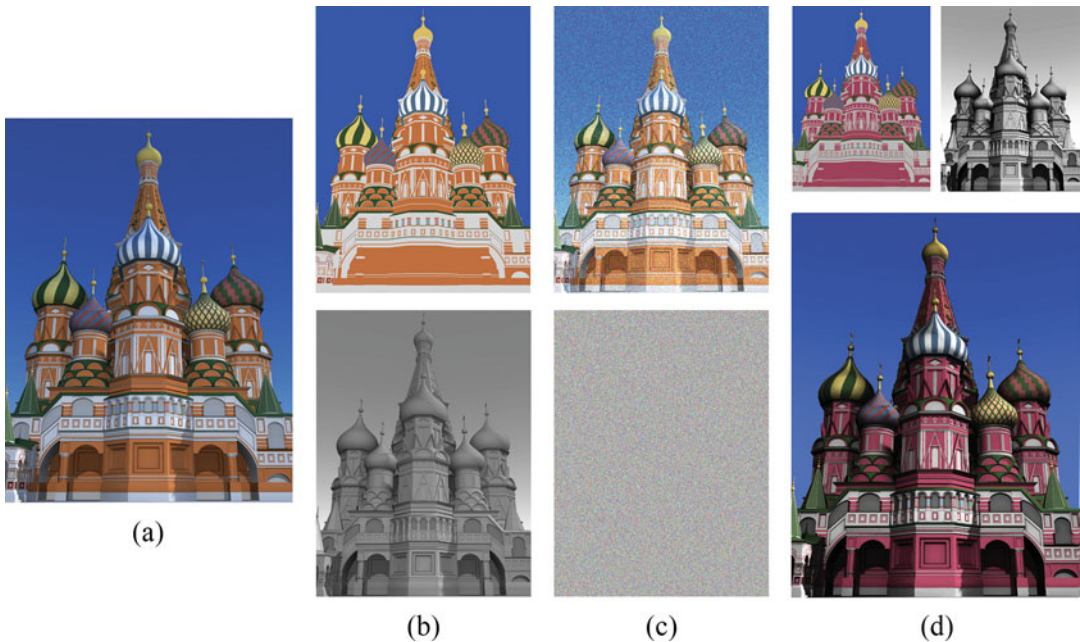
The intrinsic images model firstly introduced by Barrow and Tenenbaum [1], is a simple yet accurate way for inversely modeling the

appearance of a scene without the need to explicitly define the geometry or the illumination. Instead, it states that an image $\mathbf{I}$ captured with a conventional photographic camera can be represented as the pixel-wise multiplication of two images, as shown in Fig. 1b, where x is a pixel in the image, and $c \in \{R, G, B\}$ is the color channel or spectral band. The reflectance $\mathbf{R}$, or albedo, layer describes the diffuse color of the material and is invariant to illumination. The illumination layer $\mathbf{S}$ contains information about shading and shadows, and integrates the interaction between geometry and incoming light, usually containing also inter-reflections.

$$\mathbf{I}_c(x) = \mathbf{R}_c(x) \times \mathbf{S}_c(x) \quad (1)$$

In order to be physically accurate, this decomposition requires that the materials that compound the scene present Lambertian or diffuse reflection. That is, the reflected light (more technically, the surface luminance intensity) will be the same regardless of the observer's position, only obeying Lambert's cosine law. A huge number of materials in the real-world follow this property such as matte wall painting, unpolished wood, or freshly fallen snow. However, Lambertian reflections from polished surfaces are frequently accompanied by a specular or glossy reflection. The effect is maximally observed at the perfect reflection angle and falls off sharply. Several analytic models account for this effect such as Phong or Cook-Torrance. In the context of intrinsic decomposition, the specular highlights are usually accounted for as part of the shading layer, or less frequently, as an extra layer which is added to the initial ones, thus increasing the number of unknowns in the system.

Obtaining these components from a single image is a major challenge as for each known value of the input image there are two unknowns, therefore, infinite number of combinations that can satisfy Eq. 1 (e.g., Fig. 1d). To make the problem tractable, existing methods impose additional constraints or make certain assumptions about the components. In the following, the most common approaches are reviewed.



Intrinsic Imaging, Fig. 1 Intrinsic decomposition ambiguity (a–c) (inspired by [3]), and example of editing (d). (a) Input image **I**. (b) Decomposition into reflectance **R** (top) and shading image **S** (bottom). (c) Example of

ambiguity in the decomposition, the product of these two images also yields the same input image **I**. (d) Editing example: both layers have been modified separately and combined to alter the appearance of the input image

Priors and Methods to Constrain the Problem

Priors and constraints are some statistical tools and methods which are used to encode existing beliefs about a problem. In the intrinsic decomposition problem, several priors have been developed to help disambiguate the decomposition along with several assumptions and ad hoc user interfaces. From a general perspective, the type of constraints can be classified into: local constraints, such as those which operate at local neighborhoods around image pixels; global constraints, which extend the notion of neighborhood to any pixel in the image; and priors about the distribution of the values of each component—the shading, the reflectance, or the illumination. Besides, another type of classification can be done by analyzing the source of the knowledge which originated such constraints. This information might come: automatically from existing theories of the human vision and texture perception, defined by the user, or directly learned from

the data. In the following, the latter classification is developed in detail (see the state-of-the-art report of Bonneel et al. [3] for a more thorough evaluation with examples).

Automatic Priors

There exist a number of automatic priors that affect different aspects of the decomposition:

Shading: The initial techniques that addressed the intrinsic decomposition problem relied on theories of the human vision which relate local changes of intensity with either changes in reflectance or shading. The main theory was Retinex, which stated that the perceived color of an object remains constant as long as the reflectance ratios do not change, even if the light color changes. Inspired by it, Horn [8] implemented this theory in 2D images and presented a method to obtain the decomposition from black and white images assuming that sharp reflectance changes cause intensity discontinuities, while shading is a smooth

function. These types of constraints were implemented by thresholding local pixel derivatives to distinguish a change in reflectance from a change in shading.

Reflectance: The extension of Horn's work to colored images was done by Funt et al. [5] who associated changes in reflectance with local changes in chromaticity, and enforced integrability of the shading at these points like previous methods did. Propagating the shading values to their neighboring pixels was done by diffusion, solving the subsequent Poisson equation with a Fourier transformation. Later, several works extended this idea globally by assuming that the amount of different reflectances is sparse and performing color clustering within close-form or probabilistic frameworks [2, 6].

Texture: Textured areas provide a very strong cue about regions of the image which should have the same reflectance values. The challenge here lies in finding texture descriptors and similarity metrics among them which are invariant to changes in illumination and geometry. In an optimization framework, these constraints are imposed over both local and non-local areas of the image [12].

Lighting: The type of light that illuminates the scene is critical for its appearance. In general, most of the methods assume that the images are white-balanced, and illuminated by a single light source. This can be further constrained assuming that the scene is illuminated by a narrow-band sensor such as Planckian light or Daylight. In that case, the reflectance of a single object forms a line in log-chromaticity space. This has been implemented following the color lines model [10] or entropy minimization [7].

User Defined Constraints

Another set of techniques rely on automatic priors and user intervention. The first method to incorporate this kind of input into an optimization framework was Bousseau et al. [4]. They assumed that local reflectance variations lie in a 2D plane in RGB space, not containing black, and introduced three types of strokes with different semantics:

constant-reflectance, constant-illumination, and fixed-illumination. The result of this set of methods is pretty compelling, although creating the appropriate strokes for each image may be challenging for unskilled users.

Learning from Data

Finally, a well-known strategy to solve this problem consists of learning the constraints directly from the data. The first method using this kind of approach [11] used machine learning to classify the derivatives of the image as produced either by illumination or albedo. Ten classifiers were trained with a set of computer-generated images containing only reflectance or illumination components. Due to the high ambiguity, single classifiers were not enough, and they required to combine them with probabilistic propagation methods such as Markov Random Fields. As in many other fields in computer vision, deep learning methods have also marked a before and an after in the way this problem has been tackled. The leading strategies nowadays combine fully-supervised learning, using human-annotated and rendered datasets, with self-supervised learning, using multiple images of the same scene with varying illumination [9].

Applications

Having the intrinsic components is useful for any task that would require some knowledge about lighting or materials of the scene. For example, changing the materials of a scene is simplified to edit the reflectance layer; segmentation is improved if the reflectance layer is used instead of the original input image; or stylization and relighting can be done more accurately by optimizing the shading and reflectance components separately. An example of such edits is shown in Fig. 1d.

Cross-References

► [Rendering Natural Phenomena](#)

References

1. Barrow, H.G., Tenenbaum, J.M.: Recovering intrinsic scene characteristics from images. In: *Proceeding of the Computer Vision Systems*, New York (1972)
2. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. *ACM Trans. Graphics (Proc. SIGGRAPH)*. **33**(159), 1–12 (2014)
3. Bonneel, N., Kovacs, B., Paris, S., Bala, K.: Intrinsic decompositions for image editing. *Comput. Graphics Forum (Eurographics State of the Art Reports 2017)*. **36**(2) (2017)
4. Bousseau, A., Paris, S., Durand, F.: User-assisted intrinsic images. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*. **28**(130), 1–10 (2009)
5. Funt, B.V., Drew, M.S., Brockington, M.: Recovering shading from color images. In: *Proceeding of the Second European Conference on Computer Vision*. Springer, Heidelberg (1992)
6. Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic images by clustering. *Comput. Graphics Forum (Proc. EGSR)*. **31**(4), 1415–1424 (2012)
7. Graham, D., Finlayson, M.S.D., Lu, C.: Intrinsic images by entropy minimization. In: *Proceeding of European Conference on Computer Vision*. Springer, Berlin (2004)
8. Horn, B. K.: Determining lightness from an image. *Comput. Graphics Image Process*. **3**(4), 277–299 (1974)
9. Li, Z., Snavely, N.: CGIntrinsics: better intrinsic image decomposition through physically-based rendering. In: *Proceeding of European Conference on Computer Vision*, Munich, Germany (2018)
10. Omer, I., Werman, M.: Color lines: image specific color representation. In: *Proceeding of the Computer Vision and Pattern Recognition*. IEEE, Washington, DC, USA (2004)
11. Tappen, M., Freeman, W., Adelson, E.: Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(9), 1367–1374 (2005)
12. Zhao, Q., Tan, P., Dai, Q., Shen, L., Wu, E., Lin, S.: A closed-form solution to retinex with nonlocal texture constraints. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(7), 1437–1444 (2012)

Intrinsically Photosensitive Retinal Ganglion Cells

► [Melanopsin Retinal Ganglion Cells](#)

Iodine Lamps

► [Halogen Lamp](#)

IoT

► [Internet of Things and LEDs](#)

ipRGCs

► [Melanopsin Retinal Ganglion Cells](#)

Iridescence (Goniochromism)

Richard J. D. Tilley

Queen's Buildings, Cardiff University, Cardiff, UK

Synonyms

[Goniochromatism](#); [Irisation](#); [Labradorescence](#); [Opalescence](#); [Pearlescence](#); [Polychromatism](#); [Schiller](#) (Note: these are not *exact* synonyms: they usually apply to specific examples of iridescence)

Definition

Iridescent is the adjective used to describe objects or surfaces that show bright, metallic-looking, shining colors that change noticeably as the position of the observer changes. Iridescence is the noun that describes the colors that arise from an iridescent object.

Description and Examples

The root cause of iridescence lies in the (coherent) scattering of white incident light by transparent or semitransparent surface and subsurface structures, as was stated by Robert Hooke in his description of iridescent peacock feathers [1]. He also gave a technique for distinguishing iridescent colors from those that are due to pigmentation – when the color is wetted, it will disappear if its origin is

refraction and reflection. For color to be observed, the substructures have to vary on a size scale equivalent to about half the wavelength of visible light. Moreover, iridescent reflections are often strongly polarized.

Among inanimate examples, colors of a soap film or films of oil on water can be thought of as archetypal [2]. A number of minerals such as labradorite and opal show iridescence and are used as gemstones. Materials that oxidize to produce thin transparent films overlaying a dark background also show strong iridescence, typified by silicon carbide (SiC) and bornite (Cu_5FeS_4) also called peacock ore. Dichroic filters, which are made by coating glass discs with multilayers of oxides, show iridescent colors. Dichroic glass, made in the same way, is used by artists. Iridescent gift bags and tags are commonplace, and iridescence is utilized in car paints, cosmetics, and other decorative and artistic objects. The iridescent reflected colors from CD and DVD storage discs and holograms used as security labels are familiar (Fig. 1a–c) [3–5].

Living organisms display an enormous variety of iridescent colors [6–11]. Many creatures are able to see near-ultraviolet radiation, and iridescence in the ultraviolet is of importance in the wild even though it is undetected by humans. Thin-film iridescence similar to that in soap films can be seen on close examination of the transparent wings of flies, bees and wasps, beetles, and many other flying insects. More obviously iridescent coloring is characteristic of many groups of birds (feathers), beetles (exoskeleton), and butterflies (wing scales), although iridescent colors are not restricted to these groups. In these cases the visual effect is often enhanced by backing the iridescent structures with a melanin-related black pigment so that stray reflections do not diminish the vividness of the perceived colors. Pearls and a number of shells show iridescence, notably that of the New Zealand paua, *Haliotis iris* (Fig. 1d–f).

In biological areas there is great interest in explaining the evolution and adaptive advantage of iridescent coloration. In photonics there is considerable effort expended in mimicking natural iridescence by artificially copying biological examples.

Mechanisms

The interaction of light with an iridescent surface alters the phases of the various components of the scattered radiation, allowing constructive and destructive interference to occur. Because phase shift is sensitive to the optical path through the membranes intercepted, the colors observed change dramatically with variation in viewing angle. The mechanism of color production is often described in terms of reflection or diffraction, and although such a clear-cut division is somewhat artificial, this provides a starting point for discussion.

Reflection

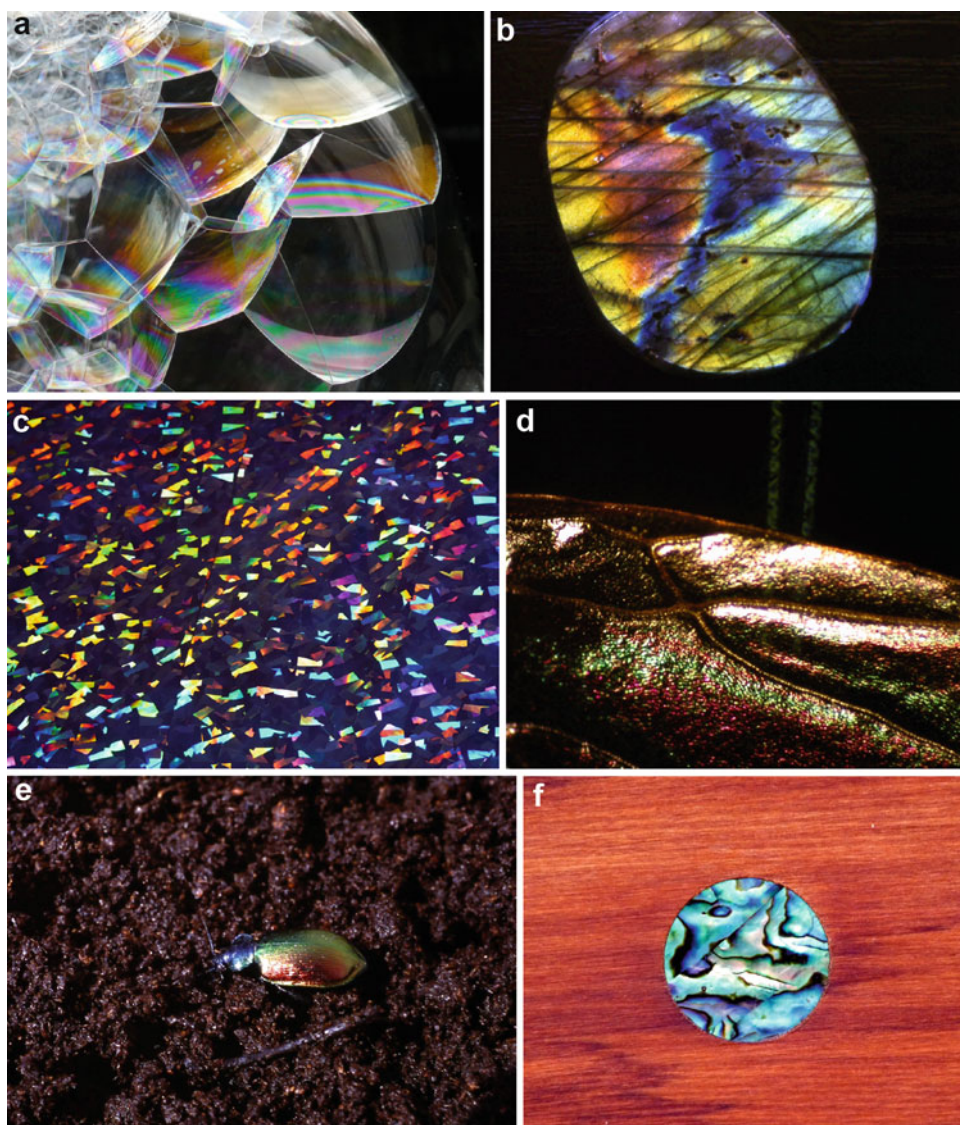
The simplest model for iridescence is the reflection of white light from a single transparent thin film, the cause of color in soap films and transparent insect wings (Fig. 1a, d). The phase shift that leads to interference and hence color production is between, in the simplest case, that part of the incident beam that is reflected off the top surface and that part that is transmitted through the film and reflected off the bottom surface (Fig. 2). At *normal incidence*, it is found that a monochromatic incident beam of wavelength λ reflected back from a thin film of refractive index n and thickness d in air or on a substrate of lower refractive index than the film will show constructive interference and so appear *bright* when:

$$2nd = (m + \frac{1}{2})\lambda \quad (m = 0, 1, 2, 3 \dots)$$

When the light is incident at a small angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos \theta = (m + \frac{1}{2})\lambda$$

If the film is on a substrate of higher refractive index, the reflected light will show destructive interference and appear dark under the same conditions. Similarly, light reflected from a film in air



Iridescence (Goniochromism), Fig. 1 Iridescence (a) soap film; (b) labradorite; (c) wrapping paper; (d) ant wing; (e) scarab beetle; (f) paua shell decoration. (Photographs R J D Tilley)

or on a substrate of lower refractive index than the film will show destructive interference and so appear *dark* when:

$$2nd = m\lambda \quad (m = 1, 2, 3 \dots)$$

If the film is on a substrate of higher refractive index, the reflected light will show constructive interference and appear bright under the same conditions. When the light is incident at a small

angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos \theta = m\lambda$$

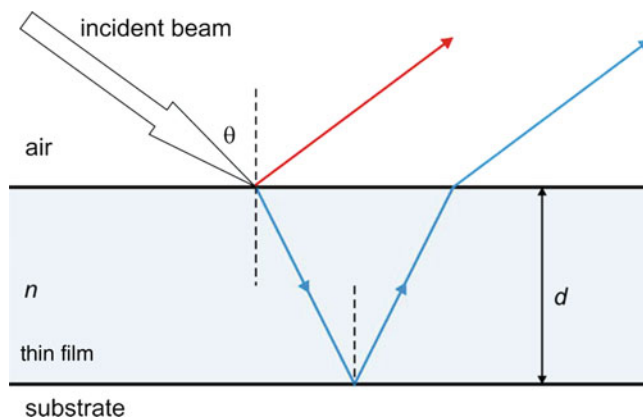
Note that the equations for viewing at an angle do not take polarization into account and must be employed with caution.

To determine the appearance of a film illuminated with white light, the reflectivity has to be

Iridescence

(Goniochromism),

Fig. 2 Production of thin-film iridescent colors is due to the phase difference between rays reflected from top (*red*) and bottom (*blue*) surfaces



summed across all of the wavelengths in the visible. The result is that some colors will be suppressed because that particular wavelength will correspond to destructive interference, while those that correspond to constructive interference will be augmented, so that the film will appear colored. The colors seen depend upon film thickness and cycle through several series or *orders*. The $\cos \theta$ angle dependence given above indicates that as the viewing angle moves away from perpendicular, the color observed moves toward shorter wavelengths: a *blueshift*. Because the color observed depends upon the film thickness and angle of observation, in the case of unstable films such as soap bubbles, the colors swirl and change as the film thins and flexes.

Many decorative films and paints are made by using thin-film iridescence. On bags and gift tags, this is often a thin polymer film, sometimes containing a dye, overlaying a reflective substrate (Fig. 1c). Metallic car paints often contain flakes of a mineral such as mica overlaid with a thin film of a transparent oxide such as TiO_2 .

A similar effect, but generally with greater brightness, is found if the reflecting object is composed of multiple layers (often called multilayer stacks or distributed Bragg reflectors) consisting of alternating lamellae of higher and lower refractive index. The mineral labradorite (Fig. 1b) contains alternating lamellae of feldspars of slightly different constitutions, thus giving lamellae with different refractive indices. The iridescent colors of many scarab beetles (Fig. 1e) are due to a multilayer structure consisting of chitin and air

layers, while in paua shells (Fig. 1f), the color arises in layers of nacre, a composite of aragonite (CaCO_3) plates and proteins. The reflectivity of multilayers depends upon the thickness of the layers, their refractive indices, and the angle of view. Relatively simple formulas have been derived for the reflectivity of stacks where two layer types alternate, each of which is $\lambda/4$ thick (called a quarter-wave stack), viewed at normal incidence. The peak reflectivity of such a stack with refractive indices n_1 and n_2 and of thicknesses d_1 and d_2 occurs at a wavelength given by $2(n_1d_1 + n_2d_2)$. The optical behavior of more complicated arrangements, such as those found in dichroic interference filters and dielectric mirrors, can be calculated using computer software.

There are a number of differences between the colors exhibited by multilayers compared to those from single thin films. In particular, the blueshift observed as the observation angle moves away from normal to the stack is much less pronounced compared to single films.

Diffraction

Colors arise when white light is diffracted by objects with a dimension similar to that of visible wavelengths. For example, a single small hole, disc, speck, or droplet gives rise to diffracted beams in the form of nested cones with maxima and minima given by:

$$\sin \theta = m\lambda/d$$

Iridescence**(Goniochromism),**

Fig. 3 Irisation shown by high clouds. (Photograph R J D Tilley)



where d is the diameter of the object and m takes values of 0, the straight through beam; 1.220, first minimum (dark ring); 1.635, first bright ring (called the first-order diffracted beam); 2.333, second minimum (second dark ring); 2.679, second bright ring (second-order diffracted beam); and so on. For orders other than 0, the angle of scattering is wavelength dependent so that each diffracted cone will consist of a spectrum, with violet diffracted least and red the most.

A random distribution of scattering centers of differing sizes will not show iridescence as the orders overlap and cancel. However, if the collection is made up of similar-sized objects, iridescence can result. This occurs on rare occasions when high clouds are made up of uniformly sized droplets and viewed at an angle of approximately $10\text{--}45^\circ$ from the direction of the sun, a phenomenon called *irisation* (Fig. 3). The fact that strong iridescent colors are only rarely seen is likely to be due to the necessity of having uniform drop sizes with diameters of the order of 0.003 mm. Similarly, iridescence arising from small evenly sized spots and droplets that form on mirrors can sometimes be seen.

A surface which contains a series of equally spaced parallel grooves or slits with a spacing of about the wavelength of visible light forms a *one-dimensional diffraction grating*. Light incident upon such a grating of spacing d (either in transmission or reflection) will be scattered strongly in directions given by the *grating equation*:

$$d (\sin \theta_i + \sin \theta_m) = m\lambda$$

where θ_i and θ_m are the angles of incidence and diffraction, respectively, of the order m and θ_m is positive when the diffracted and incident beam lie on the same side of the grating normal and negative when they lie on opposite sides of the normal. Far from the grating, the (Fraunhofer) diffraction pattern will consist of a row of bright spots called *principal maxima*. The spacing of these spots is proportional to $1/d$ and lies along a line perpendicular to the grating grooves. They are separated by much smaller secondary maxima which will not concern us here as these do not give rise to strong iridescent effects. The zero-order, $m = 0$, diffracted beam is equivalent to an undeviated beam (transmission) or specular reflection (reflection). For the other orders, the angle of scattering, θ_m , is wavelength dependent so that white light will be spread out into a spectrum, with violet diffracted least and red the most (Fig. 4). The spectra produced by such gratings are intense, giving rise to marked iridescence (Fig. 1c).

Scattering centers can be distributed across a surface to form a *two-dimensional* grating. Such a grating can consist of an array of apertures arranged in an ordered pattern in an opaque screen, an ordered set of pits or protuberances arranged on a plane, and so on, most easily visualized for a rectangular array (Fig. 5a). Each row on the pattern acts as a one-dimensional grating,

Iridescence**(Goniochromism),**

Fig. 4 Iridescent color formation by a one-dimensional reflection diffraction grating

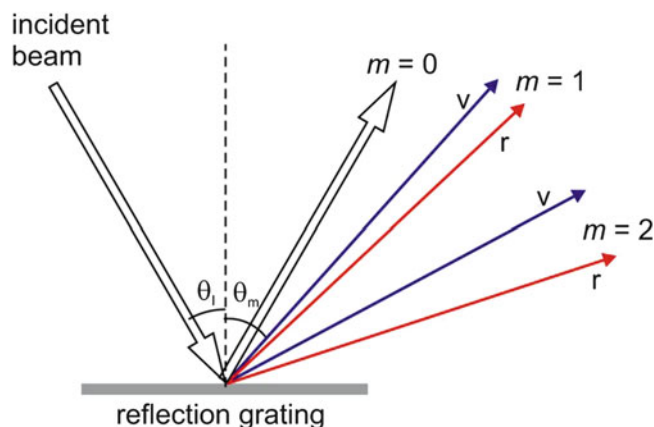
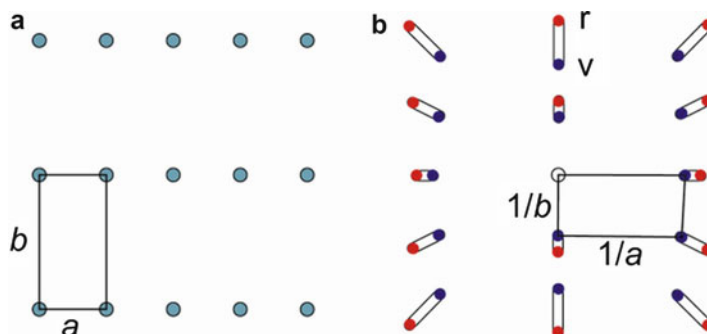
**Iridescence****(Goniochromism),**

Fig. 5 Iridescence from a two-dimensional diffraction grating: (a) grating can be envisaged as intersecting one-dimensional gratings; (b) diffraction pattern, with violet diffracted least and red most



and the diffraction pattern may be understood intuitively as arising from intersecting one-dimensional gratings. Each row will give rise to principal maxima (described by the grating equation) running normal to the row selected, separated by weak intermediary subsidiary spots that do not give rise to strong iridescence. A complete pattern will only be seen if the conditions of the grating equation are fulfilled for every row simultaneously. For a rectangular grating illuminated by a beam normal to the grating, the conditions for diffraction are:

$$a \sin \theta_m = m \lambda \text{ (grating spacing } a)$$

$$b \sin \theta_p = p \lambda \text{ (grating spacing } b)$$

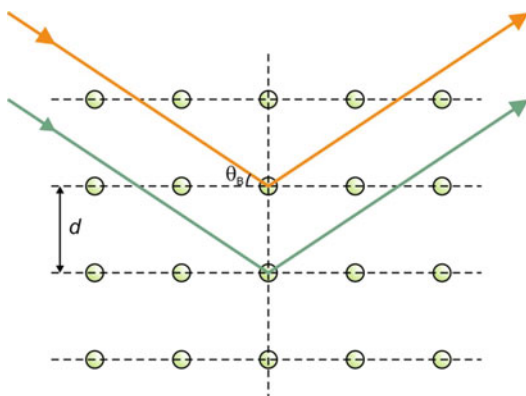
(When this condition is met, all the other rows will be perpendicular to the incident beam.) Each diffracted order will now be specified by two indices m and p , so that:

$$d_{mp} \sin \theta_{mp} = mp \lambda$$

where d_{mp} is the spacing of the grating elements along the row giving rise to the diffraction pattern row containing the order mp . As before, the zero-order beam corresponds to undeviated transmission or specular reflection and is not colored. The diffraction orders, each forming a spectrum with violet diffracted least and red the most, will now be arranged on a grid with symmetry that matches that of the grating (Fig. 5b).

Many iridescent gift tags and bags show two-dimensional diffraction effects. The surface of these consists of a thin plastic film, often overlaying a reflecting layer. The polymer surface is embossed with several sets of grooves. Two sets at right angles give rise to square or rectangular diffraction effects, while three sets at 60° give rise to hexagons and star diffraction patterns.

Natural and artificial three-dimensional ordered crystal-like arrays of scattering centers, also called *photonic crystals*, are also known [12–14].



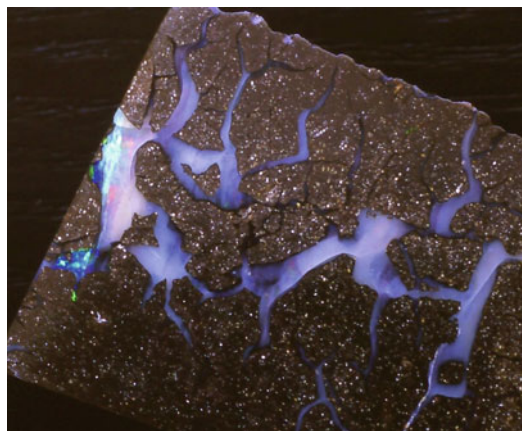
Iridescence (Goniochromism), Fig. 6 Diffraction by three-dimensional array of scattering centers

In this case diffraction is constrained by the spacing of the scattering centers in three dimensions. The condition for strong diffraction to occur is given by Bragg's law:

$$m\lambda = 2d \sin \theta_B$$

where m is the order of diffraction, d is the distance between the planes of scattering centers giving rise to the diffracted beam, and θ_B is the angle between the incident beam and the scattering planes (Fig. 6). As in previous examples, the position of all diffracted beams except for the zero order is a function of wavelength, and each diffracted order will be pulled out into a spectrum with violet diffracted least and red the most.

The most familiar natural object that displays such iridescence is the gemstone precious opal (Fig. 7). This consists of arrays of ordered silica spheres embedded in a matrix of amorphous silica. As the opal is tilted, the color observed moves to shorter wavelengths giving a blueshift. Artificial opals, also called colloidal opals, are made by sedimenting evenly sized polymer spheres. Under controlled conditions, these will pack into close-packed arrays as in natural opal and show typical opaline iridescence. If the spaces between the spheres are infiltrated with an inorganic material, typically an oxide such as TiO_2 or ZrO_2 , and the polymer subsequently dissolved, an array of close-packed shells forms. These show similar iridescence to natural opals.



Iridescence (Goniochromism), Fig. 7 Iridescent vein of opal in ironstone. (Photograph R J D Tilley)

If a structure forms in which the layers are composed of birefringent molecules, polarization of the scattered beam may occur. When the orientation of the molecules is such that there is an approximately constant rotation angle from one layer to the other, the diffracted light is circularly polarized. The color observed is given by Bragg's law, in which the pitch of the spiral is equivalent to d and the observed color changes with viewing angle as in the case of opals. This is found in a number of beetles and is exploited in liquid crystal thermometers.

Holograms

Iridescent colors that change hue and content dramatically with both viewing angle and viewing distance are produced by the holograms commonly seen as security tags on credit cards, banknotes, and many novelty gift items such as bookmarks, when viewed in white light. These are called rainbow holograms, Benton holograms, or white light transfer holograms.

Holograms are essentially interference patterns formed by the superposition of coherent light from a reference beam, and the same light after it has been scattered by an object. The "reconstructed" image is observed when the eye receives light from the reference beam alone that has interacted with the hologram. Holograms are

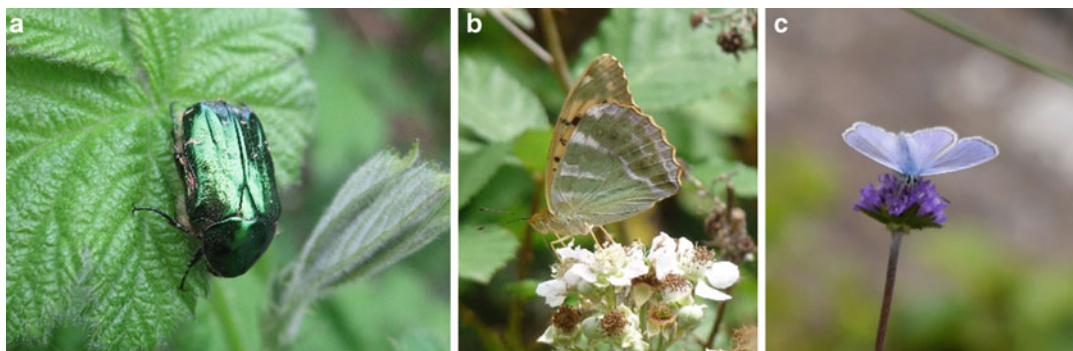
able to record both phase and amplitude information, so that the image contains considerable information content not normally inherent in an orthodox interference pattern. When conventional holograms are viewed in white light, each different wavelength produces its own distorted image of the object. These overlap and a satisfactory image is not observed. Rainbow holograms are produced by a sophisticated technique that allows undistorted colored images of the object to be viewed sequentially as the hologram is tilted. Each image of a single color appears as if viewed through a narrow slit. As the angle of observation changes, the virtual slit giving rise to the image changes, giving rise to different colors and different images.

Rainbow holograms are embossed on a thin plastic sheet by conventional embossing techniques. A master hologram is initially created, and this is copied to produce numerous metal “stamper shims” that are used in the embossing process. The interference pattern is recorded as complex series of ridges and whorls in the transparent film. To create an image, white light passes through the film containing the interference pattern and is reflected back from the backing to pass through the hologram again before reaching the eye. Passage through the holographic structure imposes a phase change on the components of the light which is interpreted as described.

Suppressed Iridescence

Many creatures, especially insects and birds, show shining metallic reflections with all of the characteristics of iridescence except that there is no significant change of hue with viewing angle, as if the iridescent color changing ability has been suppressed. These non-iridescent or suppressed iridescent colors arise in similar ways to iridescence, from coherent scattering of incident white light. The change in the optical effect is due to a change in the nanostructural organization of the scattering materials. For example, ordered high reflectivity stacks give rise to noticeable iridescence. As the stacks become less well ordered, due to refractive index variation or an uneven distribution of layers, the single angle-dependent color reflected is replaced by a shining single color that is fairly constant over a wide range of angles (Fig. 8a). If the layers in the stack are very disordered, omnidirectional metallic silver reflection is produced (Fig. 8b).

Similar complexity modifies the diffraction characteristics of diffraction grating iridescence. As an ordered grating becomes replaced by structures of a quasicrystalline nature, made of interconnected sheets, pillars, tubes, or globes of chitin or β -keratin, the characteristic angular-dependent iridescence may be replaced by a single shining color that does not change significantly with viewing angle (Fig. 8c). In many of these



Iridescence (Goniochromism), Fig. 8 Suppressed iridescence: (a) rose chafer, *Cetonia aurata* (Photograph Ralph Hobbs, 28 April 2010, Lewes, E. Sussex., with permission), (b) silver markings on the underside of the

silver-washed fritillary, *Argynnis paphia* (Photograph R J D Tilley), (c) common blue butterfly, *Polyommatus icarus*. (Photograph R J D Tilley)

natural nanostructures, Fourier transform analysis shows that short-range order is important in color production.

In some cases iridescence appears to be suppressed because the scattering medium is composed of slightly misaligned microdomains. At any particular angle of incidence, only some of the microdomains reflect coherent light which is brightly colored. As the viewpoint is changed, some of these are extinguished, while others shine out. The uniform non-iridescent shining color is the macroscopic effect of the microdomains winking on and off as the angle changes. This augments the suppressed iridescence in many blue butterflies of the lycaenid family (Fig. 8c).

Cross-References

► [Structural Color](#)

References

1. Hooke, R.: *Micrographia*, or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries therupon. The Royal Society, London (1665)
2. Isenberg, C.: *The Science of Soap Films and Bubbles*. Dover, New York (1992)
3. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
4. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
5. Lynch, D.K., Livingston, W.: *Colour and Light in Nature*. Cambridge University Press, Cambridge (1995)
6. Meadows, M.G., et al.: Iridescence: views from many angles. *J. R. Soc. Interface*. **6**(Suppl_2), S107–S265 (2009)
7. Vukusic, P.: Structural colour. In: Schwartz, J.A., Contescu, C.I., Putyera, K. (eds.) *Dekker Encyclopedia of Nanoscience and Nanotechnology*, vol. 5, pp. 3713–3722. Marcel Dekker, New York (2004)
8. Lee, D.: Chapter 10, Iridescent plants. In: Lee, D. (ed.) *Nature's Palette*. University of Chicago Press, Chicago (2007)
9. Parker, A.R.: 515 million years of structural colour. *J. Opt. A Pure Appl. Opt.* **2**, R15–R28 (2000)
10. Vukusic, P., Sambles, J.R.: Photonic structures in biology. *Nature*. **424**, 852–855 (2003)
11. Kinoshita, S., Yoshioka, S., Miyazaki, J.: Physics of structural colours. *Rep. Prog. Phys.* **71**, 076401 (2008)
12. Zhang, Y.S., Zhu, C., Xia, Y.: Inverse opal scaffolds and their biomedical applications. *Adv. Mater.* **29**, 1701115 (2017)
13. Goerlitzer, E.S.A., Taylor, R.N., Vogel, N.: Bioinspired photonic pigments from colloidal self-assembly. *Adv. Mater.* **30**, 1706654 (2018)
14. Hou, J., Li, M., Song, Y.: Patterned colloidal photonic crystals. *Angew. Chem. Int. Ed.* **57**, 2544–2553 (2018)

Irisesation

► [Iridescence \(Goniochromism\)](#)

Isochromatic Color Confusions

► [Pseudoisochromatic Plates](#)

Itten, Johannes

David Briggs¹ and Stephen Westland²

¹Short Courses, National Art School, Sydney, NSW, Australia

²Colour Science and Technology, University of Leeds, Leeds, UK



Johannes Itten was a Swiss expressionist painter, designer, and teacher and one of the main pedagogical forces behind the Bauhaus in its earliest phase. His 1961 book *The Art of Color* presented color theory in a simplified form that largely excluded scientific developments from the mid-nineteenth century onward. This simplified approach has permeated much subsequent teaching of color in art and design.

Itten was born in Södern-Linden (Switzerland) on 11 November 1888. He trained and practiced as a teacher in Bern before studying under the abstract painters Eugène Gilliard in Geneva (1912) and Adolf Hoelzel in Stuttgart (1913–1916). He then ran his own art school in Vienna until the director of the Bauhaus in Weimar, Walter Gropius, appointed him as one of its first teachers in 1919. Itten played a key role in the development of the “preliminary course” that would teach students the basics of material characteristics, composition, and color. However, conflict involving Itten’s ambitions, his promotion of the eastern-inspired Mazdaznan sect, and his opposition to involvement in commercially oriented design, by which Gropius hoped to validate the state-funded Bauhaus in a hostile political and economic climate, prompted Itten to leave in 1923. He subsequently taught in the Mazdaznan community in Zurich (1923–1926) before establishing an art and architecture school in Berlin (1926–1934) and directing the Advanced Vocational School for Textile Art in Krefeld (1932–1938). Itten then settled in Zurich, serving as director of the Museum and School of Applied Arts (1938–1953), the Silk Industry Vocational School (1943–1960), and the Rietberg Museum (1949–1956).

In his retirement, Itten published his main book on color theory, *The Art of Color*, in 1961 [1–3] and an account of his Bauhaus preliminary course, *Design and Form*, in 1963 [4]. Some of his earlier ideas on color had appeared in the rare hand-printed *Tagebuch* of 1930 [5] and *Die Farbe*, an exhibition catalog from 1944 [6].

Color Star, Color Circle, and Color Sphere

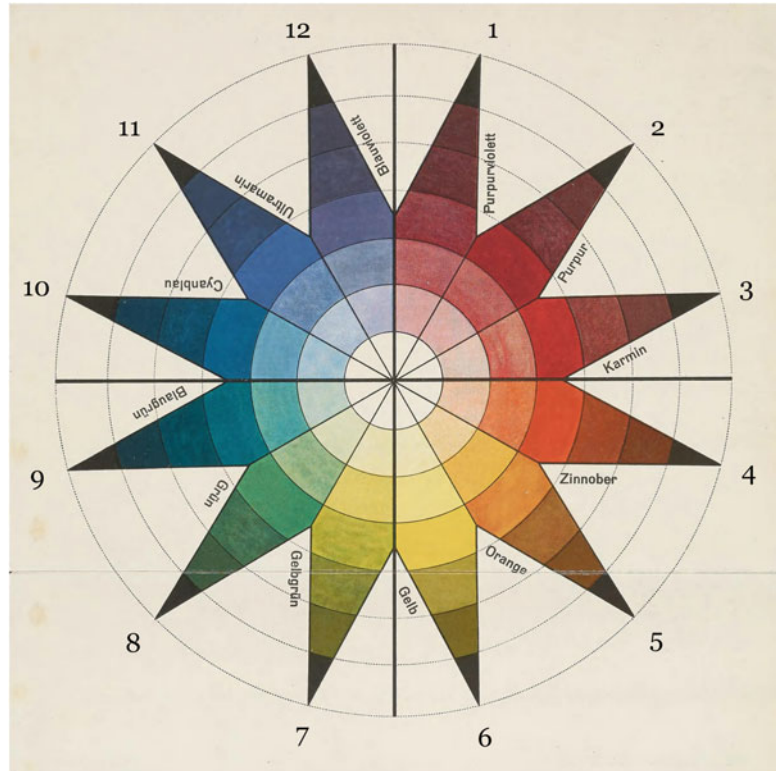
At the Bauhaus Itten taught color theory using his *Farbenkugel* (color sphere) lithograph, which included a “color star” of radiating tint and shade scales (Fig. 1). Its 12-hue scale was derived via Hoelzel from one that Bezold had proposed as being perceptually even, but was modified to align what Itten regarded as the warm-cool boundary (between yellow and yellow green) vertically. Unlike this 1921 scale, which placed yellow, “purple” (magenta?), and cyan-blue in a symmetrical triad, Itten’s 1961 color diagrams were all structured around a symmetrical triad of perceptually pure red, yellow, and blue “primary” hues, producing an unequal hue scale with larger perceptual steps in the yellow-green-blue sector. Three “secondary” hues (orange, green, and violet) and six intermediate “tertiary” hues (red-orange, yellow-orange, etc.) complete the 1961 hue scale. This change reflects Itten’s adherence to the view, widely held in science until the mid-nineteenth century, that all object colors are mixtures of red, yellow, and blue primary colors. For a three-dimensional model, Itten (1961) continued to ignore the quantitative systems devised by Munsell and Ostwald and retained a simple spherical model externally resembling the one published by Runge in 1810. This spherical model places the strongest colors of all hues on the equator, with the result that the vertical dimension does not represent lightness consistently.

Color Contrasts

Hoelzel incorporated a broad range of sources into a system of seven or eight “contrasts” of color that were central to his teaching. Itten simplified Hoelzel’s classification and language into a list of seven contrasts that is one of the most widely cited elements of his system: contrast of hue, of light and dark, of cold and warm, of complements, of saturation, of extension, and simultaneous contrast.

Itten, Johannes,

Fig. 1 Itten's color star diagram of 1921, numbered to show literal English translations of the German color names. From its position between purple-violet and crimson, "Purpur" might reasonably be interpreted as a magenta. Original diagram from Itten's "Farbenkugel" (color sphere) lithograph, included in Adler, B., *Utopia: Dokumente der Wirklichkeit*. Weimar: Utopia Verlag (1921). <https://archive.org/details/utopia00adle/page/n4/mode/2up>



- | | |
|--|-------------------------------------|
| 1 <i>Purpurviolett</i> (purple-violet) | 7 <i>Gelbgrün</i> (yellow-green) |
| 2 <i>Purpur</i> (purple) | 8 <i>Grün</i> (green) |
| 3 <i>Karmin</i> (crimson) | 9 <i>Blaugrün</i> (blue-green) |
| 4 <i>Zinnober</i> (vermillion) | 10 <i>Cyanblau</i> (cyan-blue) |
| 5 <i>Orange</i> (orange) | 11 <i>Ultramarin</i> (ultramarine) |
| 6 <i>Gelb</i> (yellow) | 12 <i>Blauviolett</i> (blue-violet) |

Subjective and "Objective" Color Harmony

Itten encouraged exploration of the color preferences of the individual, but warned that this "subjective harmony" must often be subordinated to "objective" (though scientifically unexamined) laws. "Objective" harmony required balance of the three traditional primaries, which could be obtained (following Hoelzel) from complementary pairs, equilateral and isosceles triads, and rectangular, square, and trapezoidal tetrads in his 12-hue circle, but also by tilting these shapes in any direction within his color sphere. Balance also required that the three primaries be present in a set

ratio that Hoelzel had derived ultimately from Schopenhauer [7], but which Itten misattributed to Goethe. Unbalanced, "discordant" combinations could however be used for expressive effect.

Color Expression

Itten also regarded color expression as involving objective rules and presented a system of dictionary-type meanings of colors and color combinations, in which complementary colors were expected to have opposite meanings, secondary colors were expected to combine the meanings of the primaries they "contain," and

meanings could be modified by contrast effects with surrounding colors.

References

1. Itten, J.: *The Art of Color: The Subjective Experience and Objective Rationale of Color*. Reinhold, Vienna (1961). (English translation of *Kunst der Farbe. Subjektives Erleben und objektives Erkennen als Wege der Kunst*)
2. Itten, J.: In: Birren, F. (ed.) *The Elements of Color*. Van Nostrand Reinhold Company, New York (1971)
3. Itten, J.: *The Art of Color: The Subjective Experience and Objective Rationale of Color*. Wiley, New York (1974)
4. Itten, J.: *Mein Vorkurs am Bauhaus. Gestaltungs- und Formelehre*. Otto Maier Verlag, Ravensburg (1963). (Translated into English as *Design and Form: The Basic Course at the Bauhaus*. Thames and Hudson, London (1964))
5. Itten, J.: *Tagebuch: Beiträge zu einem Kontrapunkt der bolenden Kunst*. Verlag der Itten-Schule, Berlin (1930); reissued Zurich (1962)
6. Itten, J.: *Die Farbe* (Exhibition Catalogue). Kunstgewerbemuseum, Gewerbeschule der Stadt, Zurich (1944)
7. Schopenhauer, A.: *Über das Sehn und die Farben*, 2nd edn (1854). (Translated into English by Georg Stahl as *On Vision and Colors*, p. 70. Princeton Architectural Press, New York (2010))

Further Reading

- Briggs, D.J.C.: <http://www.huevaluechroma.com/072.php#Itten> (2014). Accessed 17 Feb 2014
- Forgacs, E.: *The Bauhaus Idea and Bauhaus Politics*. Central European University Press, Budapest (1995)
- Parris, N.G.: *Adolf Hoelzel's structural and color theory and its relationship to the development of the basic course at the Bauhaus*. Ph.D. thesis, University of Pennsylvania (1979)
- Poling, C.V.: *Color theories of the Bauhaus artists*. Ph.D. thesis, Columbia University (1973)
- Wagner, C.: *Johannes Itten. Biografie in Bildern*. In: Uthemann, E.W. (eds.) *Johannes Itten: Alles in Einem – Alles im Sein*. Ostfildern-Ruit, pp. 81–96 (2003). <https://doi.org/10.11588/artdok.00001129>

Ives, Frederic Eugene

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Ives was born on February 17, 1856, in Litchfield, CT, to Hubert L. Ives and his wife Ellen. His father passed away when Frederic was still a child. He left school when he was 12 and began an apprenticeship as printer at the *Litchfield Enquirer*. Photography and engraving became his hobby. At age 18, based on his reputation and without having a formal education, he was invited to run the photography laboratory at Cornell University in Ithaca, NY, where he stayed for 4 years, a key period in regard to his inventions. While at Cornell he invented a practically valuable method of the halftone printing process for black-and-white photographs [1]. In 1879 he got married and moved to Philadelphia where he established an association with a major manufacturer of woodcut engravings who was interested in photographic reproduction methods and Ives' halftone process. Until 1880, illustrations in books, magazines, and newspapers were largely based on woodcuts as well as lithography; by the turn of the century, the industry had mostly switched over to Ives' halftone process.

ITULAB

► [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

Ives was issued a total of 70 patents in his lifetime, the earliest related to his halftone printing process and many related to trichromatic halftone printing. He revolutionized color printing in books, magazines, and newspapers in a manner similar to that for black-and-white printing. Up to that time the process in commercial use of color

printing was lithography, with the image constructed from up to a dozen handmade limestone or metal engravings used to print with differently colored inks. He regularly lectured on the subject of his inventions at Philadelphia's Franklin Institute. Circa 1890 he moved to New Jersey, outside New York. Frederic Eugene Ives died on May 27, 1937, in Philadelphia, Pennsylvania. His portrait was attached to a postage stamp celebrating his invention of the halftone process (Fig. 1) [2].

F. E. Ives' son Herbert obtained a PhD degree from Johns Hopkins University in 1908. He was also active in the field of color science, publishing in 1915 an important article on the transformation of color-matching functions between different colorimetric systems [3].

Major Accomplishments/Contributions

Ives was granted patents covering the halftone process in 1881 [4]. In 1890 he obtained a patent for his "trichromatic" color photography process



Ives, Frederic Eugene, Fig. 1 US postage stamp celebrating Ives' invention of the halftone printing process, 1996 (Copyright The U.S. Government)

Ives, Frederic Eugene, Fig. 2 Stereoscopic color image as viewed in the Kromskop, 1897. On the left are the B&W stereoscopic images for the three primary colors as used in the Kromskop. *Top right* is a digital recreation of the two stereoscopic versions of the image. On *bottom right* is an enlarged version of the image based on a photographic print made in the 1950s



[5] and in 1892 for a related camera [6]. There were two steps to Ives' halftone color printing process: (1) Color separations were obtained by making successive exposures through colored filters onto film, and halftone plates were made from these filtered exposures. (2) The three separations were printed onto paper, on top of each other in exact registration, with yellow, magenta, and blue-green inks [7]. Unlike the system proposed for color photography by J. C. Maxwell in 1855 [8] using red, green, and blue filters, Ives learned from experiments that the most accurate results were obtained when using yellow, magenta, and cyan (blue green) as primaries. This process and further developments resulted in a revolution of the technical methods of printing colored images.

In 1894 he obtained a patent for the Kromskop (Fig. 2), a stereoscopic version of his trichromatic camera [9]. It had limited success, and some years later, Ives transferred the related patents to the Eastman Kodak Company where they were used to develop the Kodachrome process. Two more significant inventions were a portable "colorimeter" and the "photometer," both patented in 1908.

The former was used to measure and define color stimuli in terms of trichromatic designations, the latter to obtain measures for hue and intensity of lights. In 1909 and 1911 Ives was granted patents for the Tripak paper color film, used in the Tripak color camera [10].

References

1. Stulik, D.C., Kaplan, A.: Halftone. Getty Institute, Los Angeles (2013)
2. <https://multimediaman.wordpress.com/> search entry: F E Ives. Accessed 5 May 2015
3. Ives, H.E.: The transformation of color mixture equations from one system to another. *J. Franklin Inst.* **115**, 673–701 (1915)
4. US Patents 237,664 and 245,501, 1881
5. US Patent 432,530, 1890
6. US Patent 475,084, 1892
7. Ives, F.E.: Photography in the colors of nature. *J. Franklin Inst.* **81**, 1–21 (1891)
8. Maxwell, J.C.: Experiments on colour, as perceived by the eye, with remarks on colour-blindness. *Trans. R. Soc. Edinb.* **21**(Part II), 275–298 (1855)
9. US Patent 531,040, 1894
10. US Patent 927,244, 1909

Judd, Deane Brewster

Michael H. Brill¹ and Rolf G. Kuehni²

¹Datacolor, Lawrenceville, NJ, USA

²Charlotte, NC, USA



Deane Brewster Judd (1900–1972) was an American physicist who contributed to the fields of colorimetry, color discrimination, color order, and color vision. Born in South Hadley Falls,

Massachusetts, he attended Ohio State University and Cornell University and received there a Ph.D. in physics in 1926. He was a Munsell research associate in colorimetry at the National Bureau of Standards (NBS) in Washington in 1926.

In 1927 Judd joined the NBS, where he remained until his retirement in 1969 and subsequently continued as a guest worker. He was the USA's representative in colorimetry at eight meetings of the International Commission on Illumination (CIE) from 1931 to 1967 and thereby a key force in the development of the CIE standard system of colorimetry, e.g., 1931 and 1964 standard observers, standard illuminants B and C, daylight illuminants like D65, and definition of colorimetric purity [1]. Largely responsible for coining the term “psychophysics,” he wrestled throughout his career with the relationship between color stimuli and color perception.

Colorimetric System

Judd introduced the concept of keeping luminosity and chromaticness separate in the CIE system. He was active in the colorimetric definition of color temperature and introduced the CIE colorimetric system to US industries. Together with D. L. MacAdam and G. Wyszecki, in 1964, he used principal component analysis to show that natural daylights are largely composed of three components from which daylights at any correlated color temperature can be defined (CIE method of calculating D illuminants) [2].

Color Difference

In a series of papers in the 1930s, Judd represented then-available color-scaling data first into a chromaticity diagram based on color-matching functions introduced by the Optical Society of America in 1922, then represented the resulting diagram in a Maxwell-type primary triangle, and finally embedding unit difference ellipses into the CIE chromaticity diagram. This work became the basis for the CIE (u,v) color diagram in 1960, slightly modified in the 1976 CIELUV color space. In 1939 he was instrumental in developing the NBS color difference formula. When in 1947, at the suggestion of the US National Research Council, the Optical Society of America (OSA) undertook to develop a perceptually uniform color space. Judd became its chairman and remained in that position until 1968, when D. L. MacAdam assumed the chairmanship, with results published in 1974. One of the key findings was “that strictly uniform color scales of all kinds are not homologous with Euclidean space” to which Judd proposed a solution implemented in the OSA Uniform Color Space [1].

Systematic Color Names

The perceived need for systematic naming of colors resulted in 1939 in ISCC-NBS Method of Designating Colors, based on the Munsell color system, with a revised edition published in 1955 [3, 4].

Color Constancy

Judd presented on color constancy first in 1940. He then presented with Helson and Warren, in 1952, results of meticulous work on the subject [5, 6].

Miscellaneous

At NBS Judd investigated impaired color vision, whiteness measurement of paper, opacity, color

stimulus measurement, and a flattery index for artificial light sources. In 1951 he proposed a modification of the CIE 1924 luminous efficiency function $V(\lambda)$ below 460 nm that became known as Judd-modified $V(\lambda)$, not implemented in the CIE system but used in some vision research works [7]. In addition to his positions at the CIE, Judd was president of the Optical Society of America from 1953 to 1955 and of the Inter-Society Color Council from 1940 to 1944. He was president of the Board of Trustees of the Munsell Color Foundation from 1942 to 1972.

In his homage, the International Colour Association (AIC) instituted its Judd Award, a prize that since 1975 is bestowed every 2 years to persons who have made relevant contributions in color research.

Judd was the author of one book [8] and more than 200 articles, 57 of which were reprinted as a volume by the NBS [1].

References

1. Judd, D.B.: In: MacAdam, D.L. (ed.) Contributions to Color Science. NBS Special Publication 545. U.S. Department of Commerce, Washington, DC (1979)
2. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
3. Judd, D.B., Kelly, K.L.: Method of designating colors. *J. Res. NBS.* **23**, 355–366 (1939)
4. Kelly, K. L., Judd, D.B.: Color, universal Language and Dictionary of Names. U.S. Department of Commerce, Washington, DC (1955/1976)
5. Judd, D.B.: Hue, saturation, and lightness of surface colors under chromatic illumination. *J. Opt. Soc. Am.* **30**, 2–32 (1940)
6. Helson, H., Judd, D.B., Warren, M.H.: Object-color changes from daylight to incandescent filament illumination. *Illum. Eng.* **47**, 221–233 (1952)
7. Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight, CIE Proceedings, vol. 1, part 7, Stockholm, 1951
8. Judd, D. B.: Color in Business, Science and Industry. Wiley, New York (1959), 2nd and 3rd edn. with Gunter Wyszecki, 1963 and 1975

K

Katz, David

Rolf G. Kuehni
Charlotte, NC, USA

Biography



Image source: University of Rostock, Germany

Katz was born on October 1, 1884 as the seventh of eight children of his parents in Kassel, Germany, where he attended the local Realgymnasium. As a child and student, he became a talented amateur

painter, interested in color and other sensory abilities. In 1902 he began studying at the nearby University of Göttingen. There he became interested in psychology taught by G. E. Müller. In 1906 he obtained a PhD degree in psychology, physics, and philosophy, and in 1907 he assumed the position of assistant to Müller. In 1911 he was named a professor, based on his work on color phenomenology [1]. From 1914 to 1918 he was a volunteer soldier in the German Army. In 1919 he received an invitation from the University of Rostock to assume a then new professorship in psychology. In the same year he married the psychologist Rosa Heine with whom he had two sons. Over the years they both were active in a wide field of psychology, including that of children and animals and had a number of well-known students. In 1929 Katz was for a time a guest professor in Maine in the USA. Shortly after Hitler's move to power in 1933, Katz and his wife were officially retired from their positions. In the same year Katz, with the help of a colleague, moved to England, followed a year later by his wife and sons. In 1937 Katz received an invitation to fill the first professorship in psychology in Sweden at the University of Stockholm. Here he was broadly active in many fields of psychology. In 1952, already retired, he received an honorary professorship from the University of Hamburg. On February 2, 1953, he died in Stockholm. Katz produced over 200 research papers and books, most in the general field of psychology [2].

Major Accomplishments/Contributions

Beginning with his dissertation, Katz was specifically interested in the phenomenology of vision, including shape, structure, space, color, and movement, to the exclusion of related physics and psychophysics. He defined three appearance modes of colors: film colors, surface colors, and volume colors. Film colors are those experienced when looking into the black tube of a spectroscope and seeing spectral colors, having a slightly spongy appearance without having a specific location in space. Surface colors are those commonly referred to as object colors. Volume colors he defined as those of colored glasses or gelatins. An important concept in his work was the subjective gray we experience either with closed eyes in the absence of strong light or with open eyes in a completely dark room. Another one was the reduction screen, a black sheet of paper with a small hole through which a small section of a colored object could be viewed and the resulting film color appearance compared to the surface color without the screen. Katz and his students performed many experiments related to color appearance involving illumination changes, surround and shadow effects, monocular versus binocular vision, exposition time, color constancy and lack thereof, light and dark adaptation, color and depth contrast effects, and the effects of chromatic illumination, to name a few. He introduced novel concepts, translated as “pronouncedness” and “insistence,” to describe nuanced phenomenological color effects. A key subject was the problem of definition of the “true” color of an object. To measure phenomenological magnitudes, he used innovative disk mixture methodology.

In his book [1] Katz discussed his agreements and disagreements with other researchers and their findings, such as Hering and of his time K. Bühler, A. Gelb, E. Jaensch, J. von Kries, and E. G. Müller. An important psychological concept of that period was Gestalt psychology based on the idea that, as its primary formulator M. Wertheimer expressed it, “the brain is holistic” by considering in assessing a given situation the total amount of information received. Together with self-organizing tendencies, this results in “the whole is greater than its parts.” Katz published a textbook on the subject in 1944, issued in several editions and multiple languages [3].

Katz’s book of 1911 was issued in a revised and updated version as *Der Aufbau der Farbwelt* (The constellation of the world of colors) in 1930 [4]. After he moved to London a shortened, translated version was published in 1935 in England as *The world of colour* [5].

References

1. Katz, D.: Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung (Appearances of colors and how they are influenced by individual experience). Barth, Leipzig (1911)
2. www.ipprdk.uni-rostock.de, search: david katz
3. Katz, D.: Gestaltpsychologie. Schwabe, Basel (1944)
4. Katz, D.: Der Aufbau der Farbwelt (The constellation of the world of colors). Barth, Leipzig (1930)
5. Katz, D.: The world of colour (R. B. MacLeod, trans.). Kegan Paul, Trench, Trubner, London (1935)

Kohlrausch, Arnt

Rolf G. Kuehni
Charlotte, NC, USA



Portrait © Jürgen Killmann and Universität von Tübingen, 1969

Biography

Arnt Kohlrausch was born on October 30, 1884 in Hannover, Germany, a member of an extended family of scientists from the early nineteenth century to the present. His father was professor of electrical technology at the Technical College of Hannover. Arnt Kohlrausch studied medicine at the universities of Marburg, München, and Rostock, graduating in 1911. In 1918 he became lecturer in physiology at the University of Berlin. In 1926 he moved to the University of Greifswald, today in the German province of Mecklenburg. In 1928 he became chair of physiology at the University of Tübingen, where he remained until his retirement in 1951. Kohlrausch was active in several fields of physiology. Perhaps his best-known work is *Körperliche und psychische Lebenserscheinungen* (Bodily and psychological phenomena of life) [2]. He also did research in visual perception and was one of the founders of the journal *Die Farbe* (Color). A peculiarity of low-level brightness perception, the *Kohlrausch-Knick* (bend) is named after him. It refers to a bend in the curve of absolute luminance threshold versus time during dark adaptation, the bend being due to the crossover of absolute thresholds of rods and cones as a function of time. The perceptual effect on an observer is that a very dim scene suddenly appears brighter as the rod sensitivity rises enough to dominate the visual signal. In English it is known as the “rod-cone break.” Kohlrausch died on July 13, 1969 [1].

Major Accomplishments/Contributions

In 1923 Kohlrausch published an article *Über den Helligkeitsvergleich verschiedener Farben, Theoretisches und Praktisches zur heterochromen Photometrie* (Comparing brightness of different colors, theoretical and practical aspects of heterochromatic photometry) [3] in which he investigated in detail the effect of hue on perceived brightness, mentioned earlier by Helmholtz [4]. This was followed in 1935 by the article *Zur Photometrie farbiger Lichter* (Photometry of colored lights) [5]. As reported by D. B. Judd [6], in 1939, J. Urbanek and E. Ferencz introduced the term Helmholtz-Kohlrausch effect for the phenomenon, a term still in use today [7]. Figure 1 shows some examples. The six circles have luminance values identical to that of the achromatic surround. To most observers the circles appear, to various degrees, lighter than the gray of the surround.

Helmholtz used the term *glühend* (glowing) to describe the appearance of the chromatic colors compared to the achromatic surround. The strength of the effect varies by hue. Quantitative investigations of the Helmholtz-Kohlrausch effect (HKE) based on the CIE colorimetric system were performed by Wyszecki and Sanders in 1964 [8], the basis of the inclusion of the HKE in the lightness formula of the Optical Society of America Uniform Color Scales [9].

K

Kohlrausch, Arnt,

Fig. 1 Examples of the Helmholtz-Kohlrausch effect. All circles have the same colorimetric lightness values as the gray surround



References

1. Kohlrausch, A.: Körperliche und psychische Lebenserscheinungen. Kohlhammer, Stuttgart (1934)
2. https://de.wikipedia.org/wiki/Arnt_Kohlrausch. Accessed 7 May 2015
3. Kohlrausch, A.: Über den Heligkeitsvergleich verschiedener Farben. Theoretisches und Praktisches zur heterochromen Photometrie. Pflugers Arch. **200**, 210–215 (1923)
4. von Helmholtz, H.: Handbuch der physiologischen Optik, 2nd edn. Voss, Hamburg (1896)
5. Kohlrausch, A.: Zur Photometrie farbiger Lichter. Das Licht. **5**, 259–275 (1935)
6. Judd, D.B.: A new look at the measurement of light and color. Illum. Eng. **53**, 61–71 (1958)
7. Urbanek, J., Ferencz, E.: Comité d'Etudes sur la Photométrie Visuelle – Rapport du Secrétariat, 10e Session, p. 44. Scheveningen, Commission Internationale de l'Éclairage (1939)
8. Wyszecki, G., Sanders, C.L.: Correlate for lightness in terms of CIE tristimulus values. J. Opt. Soc. Am. **47**, 840–842 (1957)
9. MacAdam, D.L.: Uniform color scales. J. Opt. Soc. Am. **64**, 1691–1702 (1974)

Kries, Johannes Adolph von

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



J. v. Kries

Johannes von Kries was a German physiological psychologist, or what might now be called a neuroscientist, and student of Helmholtz. In color science, he is known as the father of chromatic adaptation models for his work on the coefficient theory of adaptation. He also made significant contributions to a variety of areas including probability theory and the structure of the retina and human visual system [1].

von Kries published profusely (in German, with few works translated to English) on topics such as blood flow in arteries, duplicity theory, chromatic adaptation, zone theory, probability theory, physiology, psychology, and history. For example, observations of the Purkinje shift led von Kries to postulate the existence of two separate visual systems, rods and cones, also referred to as duplicity theory. It is well known that Helmholtz described the trichromatic theory of color vision and participated in a famous academic dispute with Hering, the main proponent of opponent-colors theory. It is less widely recognized that von Kries was among the first proponents of the so-called zone theory of color vision that allowed for trichromatic receptors (the cones) and opponent processing of the visual information. It is little wonder that von Kries is sometimes referred to as the “greatest German disciple” of Helmholtz. Considering that Helmholtz’s students included the likes of Max Planck, Wilhelm Wien, Arthur König, A.A. Michelson, and Wilhelm Wundt, that is quite a statement indeed.

In the second half of the nineteenth century, von Kries was applying probability theory to the evaluation of the effectiveness of new drugs. He realized that the computation of probability distributions depends on the classification of symptoms and pathologies into diseases, often very subjective data. Given that the important uncertainty was determined by the experimenters’ definition of events, he developed the logical foundations of a probability theory where the subjectivity of mental representations impacts the assignment of numerical values to probabilities. Interestingly, with some distortion and misunderstanding, these ideas passed on to Keynes and formed the

core of his economic theory. In his book on the topic, von Kries developed a highly original interpretation of probability, illustrating it to be both logical and objective.

Returning to color science, von Kries' most recognized and long-lived contribution came from his works on the theory of chromatic adaptation (1902, 1905) [2, 3] in which he proposed the coefficient rule that lives on to this day as the von Kries coefficient law or simply the von Kries model of chromatic adaptation. Interestingly von Kries did not write out the mathematical formula with which he is credited; he simply stated the theory in words. In MacAdam's translation of von Kries' (1902) words:

This can be conceived in the sense that the individual components present in the organ of vision are completely independent of one another and each is fatigued or adapted exclusively according to its own function.

Perhaps it shouldn't be too surprising to realize that von Kries (1902) himself foresaw that this model was too simple to explain all chromatic adaptation phenomena. The next line after his description of what is now referred to as the von Kries model reads:

But if the real physiological equipment is considered, on which the processes are based, it is permissible to doubt whether things are so simple.

Indeed things are not so simple, yet von Kries' theory forms the basis of all effective modern models of chromatic adaptation, including that embedded in CIECAM02.

References

1. von Kries, J.: *Beitrag zur Physiologie der Gesichtsempfindungen* (Physiology of visual sensations), (1878) (Translation: MacAdam, D. L.: *Sources of Color Science*. MIT Press, Cambridge, (1970))
2. von Kries, J.: *Chromatic adaptation*, *Festschrift der Albrecht-Ludwig-Universität, (Fribourg)* (1902) (Translation: MacAdam, D.L.: *Sources of Color Science*. MIT Press, Cambridge, (1970))
3. von Kries, J.: *Die Gesichtsempfindungen*. In: *Handbuch der Physiologie des Menschen*, vol. 3, pp. 109–282. Vieweg, Braunschweig (1905)

Kruithof Curve

Wout van Bommel

van Bommel Lighting Consultant, Nuenen, The Netherlands

Definition

“Kruithof's graph,” as published in 1941 by Kruithof, that shows, for general lighting in interiors, what combinations of illuminance level and correlated color temperature result in a “pleasing” or “not pleasing” assessment by persons in that interior.

Preferred Correlated Color Temperature (CCT) for General Lighting in Interiors

Around the time of the introduction of tubular fluorescent lamps in the early 1940s, Kruithof was the first who carried out a study on preferred correlated color temperature (CCT) for interior lighting. He investigated the relationship between lighting level in an interior and the CCT of the light sources used, on the assessment of people in terms of pleasantness [1]. His study received much recognition, although many subsequent researchers got inconsistent results. It must be realized that not only the lighting level but aspects as climatic conditions and cultural background influence the preferred illuminance as well.

Lighting has not only a visual effect but also a nonvisual biological effect that influences health and well-being. The CCT of lighting plays an essential role in the effectiveness of lighting for the latter aspect. This relatively new knowledge has led to dynamic interior lighting installations where both the lighting level and the CCT gradually change in the course of the day [2]. This, together with the availability of LED light sources with a wide range of CCTs, has given rise to renewed interest in the research of Kruithof and other researchers, on preferred CCT in interior lighting.

CCT and Lighting Level

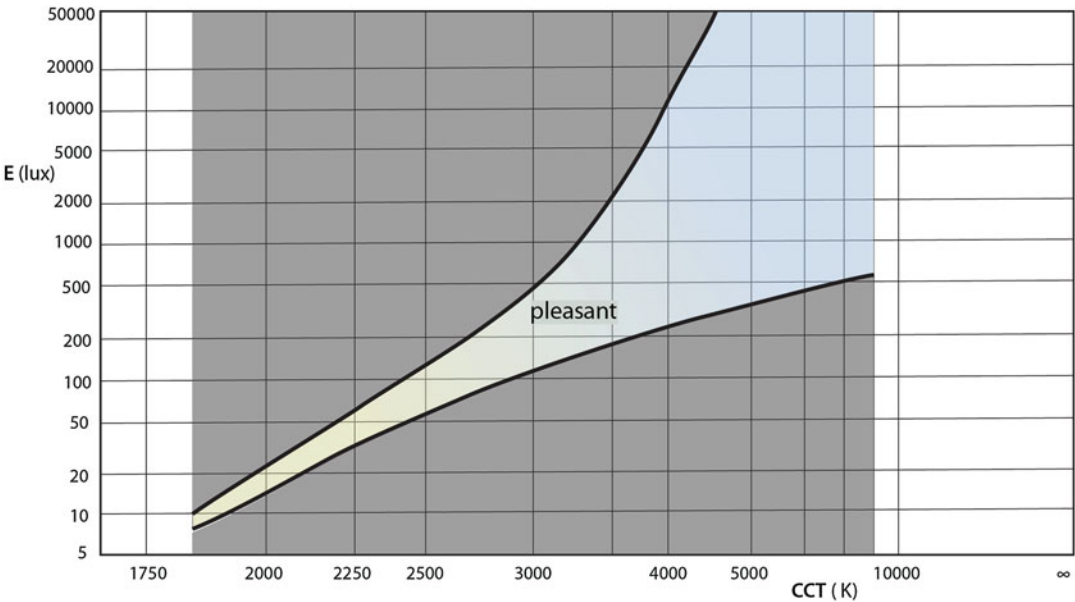
The purpose of the 1941 research of Kruithof, then one of the early developers of tubular fluorescent lamps, in fact, was the determination of the fluorescent powder composition for the production of commercial fluorescent lamps. Note that at that moment of time, manufacturers planned to bring fluorescent lamps on the market with only one or only very few types of fluorescent powder. Kruithof’s pleasantness assessment tests were carried out with observers in a room where the lighting level and the CCT of the light sources used could be changed independently. He presented the results in a graph shown in Fig. 1. Here the area in between the two curves represents illuminance level and CCT combinations that result in lighting that is judged as “pleasant.” The area above the upper curve represents the judgment “too cold” and the area under the lower curve “too dim.”

The conclusion of Kruithof is straightforward: at low illuminances, low CCT light sources

should be used and, at high illuminances, high CCT light sources. As the big market for fluorescent tubes in that era was in offices and industries where relatively high illuminances were needed, the first fluorescent tubes on the market indeed had relatively high CCTs of at least 4000 K. Another reason for these relatively high color temperatures was the fact that higher CCT lamps could be produced more efficiently.

Kruithof’s results have often been used to draw general conclusions on preferred CCTs in interior lighting design. In this context, however, it is important to note some details of his research:

- Kruithof’s graph is based on a pilot study in which only two observers participated [3].
- The research has been carried out in a laboratory setting.
- Kruithof used fluorescent lamps with different fluorescent powder compositions for the mid-range of CCTs; for the lower CCT range,



Kruithof Curve, Fig. 1 Kruithof’s graph with combinations of illuminance levels, E, and correlated color temperatures, CCTs, that are judged as “pleasant.” Logarithmic

E scale and inverted reciprocal CCT scale, as in Kruithof’s original black and white graph [1]

incandescent lamps on a dimmer; and for the higher CCT range, daylight or daylight luminance lamps. This means that the color-rendering qualities of the light sources used in his test were widely different. This may have had a considerable influence on his results.

- The observers had a short habituation time for each different situation. Recent research into the acceptance of extreme high CCT values of up to 17,000 K shows that a habituation time of some 1–2 weeks influences the acceptance of CCTs [4].
- The research has been carried out in the Netherlands (Mid-Europe). It is known that the climatic conditions and the cultural background of the observers play a role in the preference of CCT values (see next sections).

Since the publication of Kruithof, many other researchers at many different institutions have carried out investigations on the same theme. Many of them gave less straightforward conclusions [3, 5]. With the introduction of LED light sources with a wide range of CCTs for the lighting of interior spaces, the subject has received renewed interest. Again the research results are, to some degree, conflicting. Often the research results validate the statement that the use of high CCT lamps at low illuminances results in unpleasant assessments. However, low CCT lamps do sometimes result in pleasant situations when used at higher illuminances, and some studies show a general preference for the use of lower CCT lamps [6].

Based on a literature search, Fotios [7] concludes that 29 studies were published in which Kruithof's relationship was the subject of investigation. He evaluated the experimental design and quality of reporting of these studies. On this basis, only nine studies passed the test of credibility. Fotios further evaluated these nine studies resulting in his suggestion that variation in CCT (within the range of approximately 2500–6500 K) does not affect pleasing conditions. With all the knowledge gathered since Kruithof carried out his

test in 1941, the conclusion can only be that the Kruithof curves are not valid for general use.

CCT and Climatic Conditions

The predominantly prevailing climate also has an influence on the preference for color impression in interiors. In a warm climate, the preference is more toward light sources with higher CCTs, while in colder climates, preference is more for lower CCTs. These preferences are confirmed by sales analysis of the various color types of fluorescent tubes per region. With the availability of dynamic lighting installations with which CCT can be varied, sometimes the CCT during the winter and summer season is varied according to the above rule, especially in shop lighting.

CCT and Cultural Background

The cultural background plays, probably because of longtime habituation, a role in the acceptance and preference of both CCT and light levels. A typical illustration is Japan, one of the few countries in the world where fluorescent tubes have been used in domestic lighting, right from the beginning of their introduction in the 1940s. These lamps were originally only available in CCT versions of around 4000–5000 K. Today, in Japan, a preference for higher CCT lamps is seen than in other countries with similar climatic conditions where the standard in domestic lighting was incandescent lamps with relatively low CCTs.

CCT and Nonvisual Biological Effects

The eye has light-sensitive cells (intrinsic photo-sensitive retinal ganglion cells) that connect with the biological clock (suprachiasmatic nucleus) in the brain, which in turn connects with the pineal gland that controls for a part the rhythm of the hormone metabolism. Through these pathways, light has a nonvisual biological effect that influences health and well-being. The intrinsic photo-sensitive retinal ganglion cells have a high sensitivity in the blue part of the spectrum. Cool-white light of a high CCT has, therefore, a more significant nonvisual biological effect than light

with a large red component warm-white light of low CCT.

This relatively new knowledge has already led to two new developments in lighting, especially in office, industrial, school, and hospital lighting:

- Dynamic lighting installations that automatically vary the lighting level and CCT in the course of the day to get, at the right moments, biologically activating and relaxing effects
- The introduction and use of extreme high CCT lamps with good color rendering ($R_a > 80$) for use in interior lighting installations (e.g., in dynamic installations as described above)

These developments will probably also have a meaning for the preferred CCT in interiors.

Cross-References

- [Color Preference](#)
- [Human-Centric Lighting](#)
- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)

References

1. Kruithof, A.A.: Tubular luminescence lamps for general illumination. *Philips Tech. Rev.* **6**, 65–96 (1941)
2. Van Bommel, W.J.M., van den Beld, G.J.: Lighting for work, a review of visual and biological effects. *Light Res. Technol.* **36**, 255–269 (2004)
3. Van Bommel, W.J.M.: *Interior Lighting Fundamentals, Technology and Application*. Springer, Cham (2019)
4. Mills, P.R., Tomkins, S.C., Schlangen, L.J.M.: The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. *J. Circadian. Rhythm.* **5**, 2 (2007)
5. Boyce, P.R., Cuttle, C.: Effect of correlated color temperature on the perception of interiors and color discrimination performance. *Light Res. Technol.* **22**, 19–36 (1990)
6. Vienot, F., Durand, M., Mahler, E.: Kruithof's rule revisited using LED illumination. *J. Mod. Opt.* **56**, 1433–1446 (2009)
7. Fotios, S.: A revised Kruithof graph based on empirical data. *LEUKOS*. **13**(1), 3–17 (2017)

Kubelka, Paul

Michael H. Brill¹ and Michal Vik²

¹Datacolor, Lawrenceville, NJ, USA

²Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic



Paul Kubelka (1900–1954) was a Czechoslovakian chemical engineer whose many accomplishments include a theory of light absorption and scattering by a layer of paint. That theory, originally published in 1931 with Franz Munk [1], underlies much software that performs colorant recipe prediction (colorant formulation). Whereas the 1931 theory assumed that light flows in one dimension (two fluxes, upward and downward within the layer), in 1948 Kubelka derived the same equations (up to a factor of 2) assuming spherical scatter within the paint layer [2]. Later he generalized the theory to inhomogeneous layers [3].

The main contribution of these articles was a closed-form function relating the reflectance R of a layer to two constants characteristic of small particles within the layer: the absorption

coefficient K and the scattering coefficient S (both assessed in a unit thickness of the layer). If the layer is opaque, then the reflectance is a function of K/S , otherwise R depends on K and S separately, as well as on the reflectance of the material behind the layer. The Kubelka-Munk analysis also includes equations for total transmittance of a translucent layer.

Although Kubelka and Munk independently developed their analysis for paint layers, the underlying theory originated with an astronomical motivation, starting with Arthur Schuster's 1905 paper [4] relating to transmission of light through clouds.

To render Kubelka-Munk analysis useful for colorant formulation, one needed the additivity principle described by Duncan in 1940 [5]. This principle says that the total absorption coefficient K of a layer is the concentration-weighted sum of the K values of the components ($K = c_1 K_1 + c_2 K_2 + \dots$) and similarly for S ($S = c_1 S_1 + c_2 S_2 + \dots$).

Once the Kubelka-Munk theory, the additivity principle, and computer technology had emerged, colorant formulation was on its way. By 1958 Davidson and Hemmendinger introduced the analog Colorant Mixture Computer (COMIC), and this was quickly followed by dedicated digital devices, which in turn yielded to software packages that ran on general-purpose digital computers.

The above discussion places Paul Kubelka's color-science contribution in historical context. His life story offers an interesting context as well (see his 1947 autobiography <http://www.graphics.cornell.edu/~westin/pubs/kubelka-autobio.html>, accessed 8 Jan 2014).

Kubelka was born in 1900 to Austrian parents in Czechoslovakia. He was educated in the German language and attended elementary school in Kladno and secondary school in Brno and Prague. In 1918 he served half a year in the Austrian Army and began studies at the Technical University in Prague. In 1922 he passed his final examination as a chemical engineer. He then served seven months in the Czechoslovakian Army, where he soon commanded the military analytical laboratory. He also worked with L. Storch, professor of physical chemistry at the Technical University.

After leaving military service, Kubelka collaborated with Werner Mecklenburg, the well-known colloid chemist, at Verein für Chemische und Metallurgische Produktion, in Aussig (SPOLCHEMIE, in Ústí n. L.), Czechoslovakia. During this period, he worked on activated charcoal, resulting in patents and in the gas mask charcoal "G 1000." One theoretical investigation of this period earned Kubelka a doctorate of engineering in 1926.

During this time Kubelka married Margarethe Schönhöfer and had two children.

In 1928 he led both the Inorganic and Analytic laboratories in Ústí n. L. His laboratory investigated (among other things) high temperature reactions, the preparation of pigments, and activated charcoal.

In 1931 Kubelka resolved to enter an academic career. Because of his publication on absorption and capillary condensation, he was nominated docent of the University of Prague. There he investigated the absorption of vapors by silica gel, which led to an exact method of measurement of surface tension of crystals. His work interested Fritz Haber, and he expected to be nominated as professor at a German University. (It was during this year that Kubelka published the famous paper with Munk, but he does not mention this in his autobiography.)

In 1933 the situation was changed by the Nazi revolution in Germany. Kubelka refused to go to Germany and had few choices in the rest of Central Europe, so he returned to technological contributions – soon founding the company Kubelka Schuloff & Co., which eventually changed to Dr. P. Kubelka & Co. The company established a research laboratory and later a small factory. Products such as the fungicide Cuprenox were successful. References [6] and [7] are two representative patents from that era. But World War II disturbed this success and prevented the realization of other inventions. Furthermore, Kubelka was forced to change to German citizenship and then was ostracized by the Nazis.

During the postwar events of 1945, Kubelka and his wife were interned by mistake. His wife died in the camp, and he was released finally by the efforts of Czech and Jewish friends in September 1945. His German citizenship was

nullified, he was promised to regain Czechoslovak citizenship, and he accepted a position as a research chemist at the Film Company at Cesky Brod near Prague, branch factory of the Aussig Combine (SPOLCHEMIE). There he worked out a new photomechanical emulsion and reorganized the testing system. Upon deciding to go to America, he was told that the Czech authorities intended to prevent him as a specialist from leaving the country. The only place to emigrate legally was to Germany. Through the date of writing of his autobiography, he then lived in Bavaria with his children, working on the optical theory of light-scattering materials and thermodynamics of absorption and capillary condensation. In March 1947 he married Dr. Brigitte Gade.

After emigrating to Brazil in 1950, Paul Kubelka conducted further research to generalize his optical transfer theory to inhomogeneous layers [3]. He passed away in 1954 in Rio de Janeiro, where he was head of the Research Laboratory at the Brazilian Ministry of Agriculture.

References

1. Kubelka, P., Munk, F.: Ein Beitrag zur Optik der Farbanstriche. *Z. Techn. Physik*, **12**, 593–601 (1931). See also English translation by S. Westin (An article on optics of paint layers, <http://www.graphics.cornell.edu/~westin/pubs/kubelka.pdf>)
2. Kubelka, P.: New contributions to the optics of intensely light-scattering materials. Part I. *J. Opt. Soc. Am.* **38**, 448–457 (1948) [Also see errata, *ibid* p. 1067]
3. Kubelka, P.: New contributions to the optics of intensely light-scattering materials. Part II. Non-homogeneous layers. *J. Opt. Soc. Am.* **44**, 330–334 (1954)
4. Schuster, A.: Radiation through a foggy atmosphere. *Astrophys. J.* **21**, 1–22 (1905). [see <http://articles.adsabs.harvard.edu/full/1905ApJ....21....1S/0000001.000.html>]
5. Duncan, D.R.: The colour of pigment mixtures. *Proc. Phys. Soc. Lond.* **52**, 380–390 (1940). [also see his article by the same title in *J. Oil Colour Chem. Assoc.* **32**, 296–321 (1949)]
6. Kubelka, P., Srebeek, J.: Rutile pigments and process of making the same. US Patent 2,062,133 (1936)
7. Kubelka, P., Srebeek, J.: Rutile pigments. US Patent 2,062,134 (1936)

L

Lab

- [CIELAB for Color Image Encoding \(CIELAB, 8-Bit; Domain and Range, Uses\)](#)

Labradorescence

- [Iridescence \(Goniochromism\)](#)

Lambert, Johann Heinrich

Rolf G. Kuehni
NCCharlotte, USA

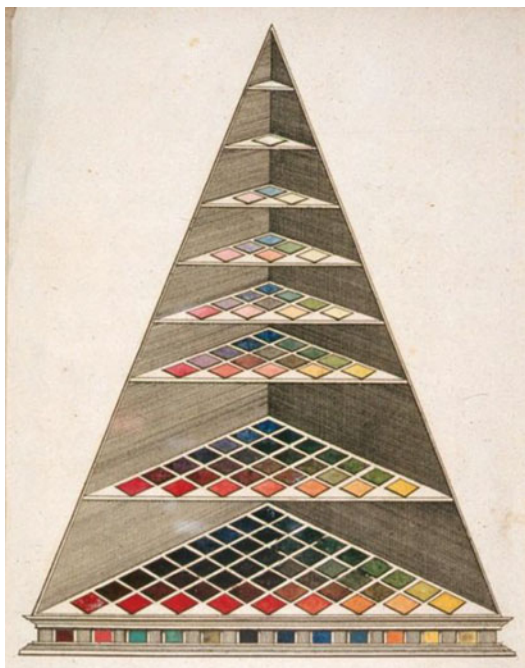
Biography

Lambert was born on August 26, 1728 in the city of Mulhouse, then an enclave of Switzerland (now part of France). He was largely self-educated, going to school only until age 12. By age 17 he assumed the job of secretary to a newspaper publisher in nearby Basel, Switzerland. He also began to work as a private tutor. At age 20 he became tutor to three boys in the family of Count Peter von Salis in Chur, Switzerland, a position he held for 10 years. There he had access to the count's large library and was able to travel widely in Europe with his charges. In 1755 he began to

publish scientific articles on a number of subjects. In 1756 he traveled with his pupils to Göttingen in Germany where he met Tobias Mayer and was elected a member of the *Königliche Gesellschaft der Wissenschaften* (Royal Society for the Sciences). In 1759 he published his work on light measurement, *Photometria* [1], introducing his mathematical formula for the law of absorption of light (Lambert's law), described non-mathematically a few years earlier by Pierre Bouguer. In 1764 he followed an invitation by the Swiss mathematician Leonhard Euler to come to Berlin where, after some initial difficulties, Frederic II appointed him to a position in the *Königlich-Preussische Akademie der Wissenschaften* (Royal Prussian Academy of Sciences). Lambert established an important position as mathematician, physicist, astronomer, and philosopher. He also had considerable interest in the art of painting. Among many other achievements, Lambert was the first to mathematically prove the irrationality of the number π . He died on September 25, 1777 in Berlin, Germany [2].

Major Accomplishments/Contributions

In 1758 Tobias Mayer presented his public lecture on a three-dimensional color order system, of which a report was published in *Göttingische Anzeigen für gelehrte Sachen* (Göttingen reports on learned matters), read by Lambert. In 1768 Lambert published an article *Mémoire sur la partie*



Lambert, Johann Heinrich, Fig. 1 Lambert's illustration of his triangular color pyramid, 1772

photométrique de l'art du peintre (Dissertation on the photometric component of the art of the painter) [3] in which he discussed the effect of light on the appearance of colored materials. Soon after and as a result of Mayer's premature death, he began work on an implementation of Mayer's conceptual double pyramidal system, with assistance of the Prussian court painter Benjamin Calau (1724–1785). The result was published in 1772 as *Beschreibung einer mit Calauischem Wachse ausgeführten Farbenpyramide* (Description of a color pyramid painted with Calau's wax) [4] containing a hand-colored abbreviated version of the conceptual color pyramid (Fig. 1) [3]. Lambert and Calau had to solve a number of practical issues, for example, they determined the relative strength of the pigments they used. When mixing the three primary paint samples, they obtained near-black colors. As a result, Lambert saw no need for the lower half of Mayer's double pyramid. The Lambert-Calau pyramid is the first three-dimensional illustrated representation of a systematically developed color solid.

References

1. Lambert, J.H.: *Photometria, sive de Mensura et Gradibus Luminis, Colorum et Umbrae*. Klett, Augsburg (1760)
2. Kraus, A.: Lambert, Johann Heinrich. In: *Neue Deutsche Biographie*, vol. 13, pp. 437–439. Duncker & Humblot GmbH, Berlin (1982)
3. Lambert, J.H.: *Mémoire sur la partie photométrique de l'art du peintre*. *Mémoires de l'Académie des Sciences de Berlin*. XXIV, 313–334 (1768)
4. Lambert, J.: H.: *Beschreibung einer mit Calauischem Wachse ausgeführten Farbenpyramide*. Haude und Spener, Berlin (1772) English translation available at www.iscc.org

Land Retinex Theory

► Retinex Theory

Lanthanide Ions

► Lanthanoid Ion Color

Lanthanides

► Lanthanoid Ion Color

Lanthanoid Ion Color

Richard J. D. Tilley
Queen's Buildings, Cardiff University, Cardiff,
UK

Synonyms

Lanthanide ions; Lanthanides; Rare earth elements; Rare earth ions; Rare earths

Definition

The lanthanoids (often designated Ln) are the 15 elements with atomic numbers 57 (lanthanum) to 71 (lutetium).

Colors, Electron Configurations, and Energy Levels

Color

Most of the lanthanoid ions exhibit rather pale characteristic colors when introduced into transparent solids or in water solutions: the most important being the Ln^{3+} state (Table 1). These colors arise from electronic transitions between the ionic ground state and energy levels derived from 4f electron configurations lying between 1.77 eV and 3.10 eV above it, giving absorption

maxima in the visible wavelength range (700–400 nm). Of more practical importance is color produced when ions excited to higher energy levels fall back to these 4f-derived levels and thence to the ground state, giving rise to characteristic visible emission spectra, which are used in many applications including fluorescent printing inks used as security markers on banknotes.

The 4f electrons in the lanthanoids are well shielded beneath an outer electron configuration ($5s^2\ 5p^6\ 6s^2$) and so are little influenced by the surrounding solid matrix, and although crystal-field effects (see ► “Transition Metal Ion Colors”) contribute to the fine structure of the electronic spectra of the lanthanoid ions, these do not have a gross effect upon the color. This implies that the most important optical properties attributed to the 4f electrons on any particular lanthanoid ion do not (usually) depend significantly upon the host structure, so that lanthanoid elements find use in phosphors, lasers, and other light-emitting devices, where a host lattice can be chosen with respect to processing conditions without significantly changing the desirable color properties of the ion.

Lanthanoid Ion Color, Table 1 Characteristic color of lanthanoid ions

Element	Ion ^a	4f occupancy ^b	Color
Lanthanum	La^{3+}	$4f^0$	Colorless
Cerium	Ce^{4+}	$4f^0$	Colorless
	Ce^{3+}	$4f^1$	Pale yellow
Praseodymium	Pr^{4+}	$4f^1$	Colorless
	Pr^{3+}	$4f^2$	Green
Neodymium	Nd^{3+}	$4f^3$	Lilac/violet
Promethium	Pm^{3+}	$4f^4$	Pink
Samarium	Sm^{3+}	$4f^5$	Pale yellow
	Sm^{2+}	$4f^6$	Red/green
Europium	Eu^{3+}	$4f^6$	Pink
	Eu^{2+}	$4f^7$	Red-brown
Gadolinium	Gd^{3+}	$4f^7$	Colorless
Terbium	Tb^{4+}	$4f^7$	Colorless
	Tb^{3+}	$4f^8$	Pale pink
Dysprosium	Dy^{3+}	$4f^9$	Pale yellow
	Dy^{2+}	$4f^{10}$	Brown
Holmium	Ho^{3+}	$4f^{10}$	Yellow
Erbium	Er^{3+}	$4f^{11}$	Pink
Thulium	Tm^{3+}	$4f^{12}$	Green
	Tm^{2+}	$4f^{13}$	Green
Ytterbium	Yb^{3+}	$4f^{13}$	Colorless
	Yb^{4+}	$4f^{14}$	Colorless
Lutetium	Lu^{3+}	$4f^{14}$	Colorless

^a Ln^{3+} is the principal ionic state encountered
^bThere is uncertainty about the exact f-orbital occupation in many compounds

Lanthanoid Free-Ion Energy Levels

The energy levels of a free lanthanoid ion are usually labeled with atomic *term symbols* derived by *Russell-Saunders (LS) coupling* (although other coupling schemes are also used in this respect). A *term* is a set of states which are very similar in energy, and the appropriate term symbol is written as ^{2S+1}L where L is a many-electron quantum number describing the total orbital angular momentum of all of the electrons surrounding the atomic nucleus and S is a many-electron quantum number representing the total electron spin. The superscript ($2S + 1$) is called the *multiplicity* of the term and is given a name: 1, singlet; 2, doublet; 3, triplet; 4, quartet; and so on. The total angular momentum quantum number L is given a letter symbol: $L = 0$, S; $L = 1$, P; $L = 2$, D; $L = 3$, F; and thereafter alphabetically, omitting J. The energies of the terms must be determined by quantum mechanical calculations, except for that of the ground state, which is given by Hund’s second

rule: the ground state is the term with the highest multiplicity and, if more than one term of the same multiplicity is present, by that with the highest L value.

The term symbol does not account for the true complexity of the energy levels of the lanthanoid ions. This arises from the interaction between the spin, S , and the orbital momentum, L , called *spin-orbit coupling*. For this the quantum number, J , is needed. It is given by:

$$J = (L + S), (L + S - 1) \dots |L - S|$$

where $|L - S|$ is the modulus (absolute value) of the quantity $L - S$. Each value of J represents a different energy level. The new quantum number is incorporated as a subscript to the term, now written $^{2S+1}L_J$ and called a *level*. It is found that a singlet term always gives rise to one level, a doublet two, a triplet three, and so on (Table 2). The energies of these levels can be sorted in terms of energy by Hund's third rule: the level with the lowest energy is that with the lowest J value if the valence shell is up to half full and that with the highest J value if more than half full. The separation between the components of the spin-orbit energy levels is of the order of 0.1–0.25 eV. (For

comparison, crystal-field splitting of these energy levels, which is due to the interaction of the f orbitals with the surrounding atoms in a solid or liquid, is about 0.01 eV.)

In the presence of a magnetic field, the spin-orbit levels are split further due to the Zeeman effect. The same is true of static electric fields, where it is called the Stark effect. In both cases, atoms or ions in a gas or free space will show an average effect because of the motion of the particles. However, in a crystal the atom and ion positions are more or less fixed, and the application of either magnetic or electric fields along certain symmetry directions will, in general, cause different degrees of splitting of the levels than the same fields applied along other symmetry directions. Zeeman splitting has been used to change the optical properties of lanthanoid ions in optomagnetic materials.

Selection Rules

Electron transitions are governed by selection rules that give the *probability* that the transition will occur. Transitions between energy levels derived purely from f orbitals are forbidden by the Laporte selection rule. However, this rule may break down for ions in compounds. The main reason for this is a degree of mixing between s , p , d , and f orbitals can occur when an ion is not located at a center of symmetry. As s , p , or d to f transitions are allowed, transitions giving rise to color are also allowed, to a degree corresponding to the amount of orbital mixing achieved. In addition, transitions are only allowed between states of the same multiplicity, called *spin-allowed transitions*. Transitions between states of differing spin can be weakly allowed and in some circumstances can also contribute to observed lanthanoid color. The weak colors exhibited by Ln^{3+} ions are primarily due to these restrictions, especially when compared to typical crystal-field colors of the 3d transition-metal ions (see ► “Transition Metal Ion Colors”).

Some Lanthanoid Absorption Colors

The $4f^0$ ions La^{3+} and Ce^{4+} and f^{14} ions Yb^{4+} and Lu^{3+} have no f – f energy levels and are colorless.

Lanthanoid Ion Color, Table 2 Ground state terms and levels of principal lanthanoid ions

Ion	Term	Levels ^a
$\text{La}^{3+} f^0$	^1S	$^1\text{S}_0$
$\text{Ce}^{3+} f^1$	^2F	$^2\text{F}_{5/2}$ $^2\text{F}_{7/2}$
$\text{Pr}^{3+} f^2$	^3H	$^3\text{H}_4$ $^3\text{H}_5$ $^3\text{H}_6$
$\text{Nd}^{3+} f^3$	^4I	$^4\text{I}_{9/2}$ $^4\text{I}_{11/2}$ $^4\text{I}_{13/2}$ $^4\text{I}_{15/2}$
$\text{Pm}^{3+} f^4$	^5I	$^5\text{I}_4$ $^5\text{I}_5$ $^5\text{I}_6$ $^5\text{I}_7$ $^5\text{I}_8$
$\text{Sm}^{3+} f^5$	^6H	$^6\text{H}_{5/2}$ $^6\text{H}_{7/2}$ $^6\text{H}_{9/2}$ $^6\text{H}_{11/2}$ $^6\text{H}_{13/2}$ $^6\text{H}_{15/2}$
$\text{Eu}^{3+} f^6$	^7F	$^7\text{F}_0$ $^7\text{F}_1$ $^7\text{F}_2$ $^7\text{F}_3$ $^7\text{F}_4$ $^7\text{F}_5$ $^7\text{F}_6$
$\text{Eu}^{2+} f^7$	^8S	$^8\text{S}_{7/2}$
$\text{Gd}^{3+} f^7$	^8S	$^8\text{S}_{7/2}$
$\text{Tb}^{3+} f^8$	^7F	$^7\text{F}_6$ $^7\text{F}_5$ $^7\text{F}_4$ $^7\text{F}_3$ $^7\text{F}_2$ $^7\text{F}_1$ $^7\text{F}_0$
$\text{Dy}^{3+} f^9$	^6H	$^6\text{H}_{15/2}$ $^6\text{H}_{13/2}$ $^6\text{H}_{11/2}$ $^6\text{H}_{9/2}$ $^6\text{H}_{7/2}$ $^6\text{H}_{5/2}$
$\text{Ho}^{3+} f^{10}$	^5I	$^5\text{I}_8$ $^5\text{I}_7$ $^5\text{I}_6$ $^5\text{I}_5$ $^5\text{I}_4$
$\text{Er}^{3+} f^{11}$	^4I	$^4\text{I}_{15/2}$ $^4\text{I}_{13/2}$ $^4\text{I}_{11/2}$ $^4\text{I}_{9/2}$
$\text{Tm}^{3+} f^{12}$	^3H	$^3\text{H}_6$ $^3\text{H}_5$ $^3\text{H}_4$
$\text{Yb}^{3+} f^{13}$	^2F	$^2\text{F}_{7/2}$ $^2\text{F}_{5/2}$
$\text{Lu}^{3+} f^{14}$	^1S	$^1\text{S}_0$

^aIn ascending energy order, ground state level **bold**

The colorless ions Gd^{3+} and Tb^{4+} have a stable $4f^7$ configuration, and there are no energy levels in the appropriate energy range to give rise to color. The same is true for the $4f^{13}$ ion Yb^{3+} .

Ce^{3+} and Eu^{2+}

The lowest energy levels of Ce^{3+} , arising from the single f electron, are $^2F_{7/2}$ and $^2F_{5/2}$. The next higher energy state for Ce^{3+} is the 5d level. Due to the interaction of the more exposed 5d electrons with the surrounding crystal structure, this is broadened into a band of energies, which also may overlap with another broadened band of energies derived from the 6s energy level (Fig. 1a). Transitions between the 5d band and 4f levels are allowed, and the colors produced by transitions of this type are intense. This transition absorbs at the violet end of the spectrum, and the absorption band often encroaches into the visible with a consequence that to the eye Ce^{3+} compounds are perceived as pale yellow.

Eu^{2+} , with a configuration $4f^7$, also has a simple energy level diagram because the energy of the state obtained by transferring an f electron to the outer 5d orbitals is lower than the other 4f energy levels (Fig. 1b). As in the case of Ce^{3+} , transitions from the ground state to the upper energy band are allowed. The energy gap is slightly smaller than in the case of Ce^{3+} , and so the absorption moves slightly deeper into the visible spectrum. Because of this, the color of Eu^{2+} compounds is described as red-brown.

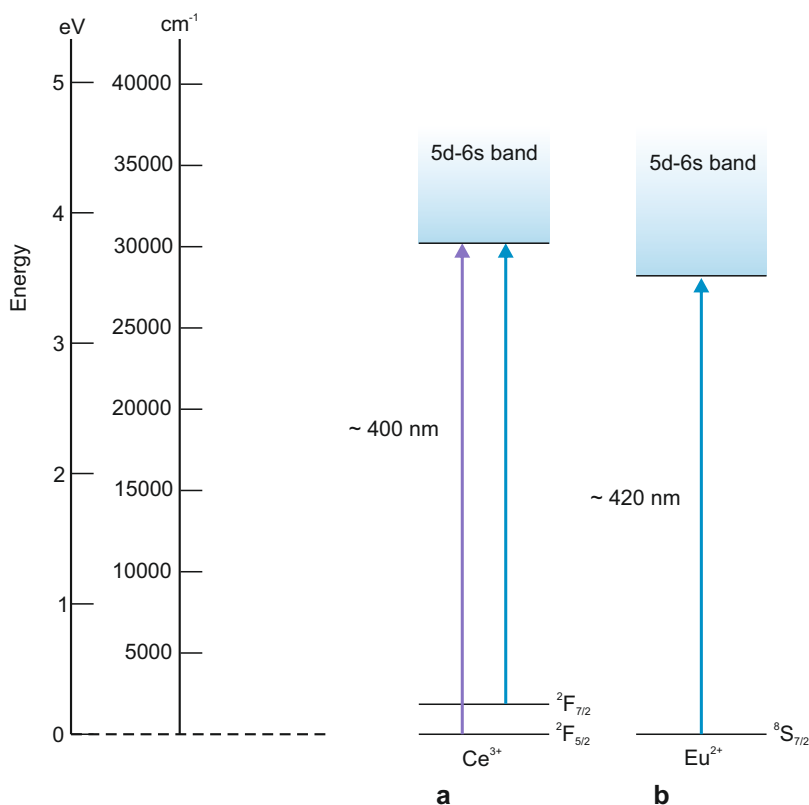
Note that as the d orbitals interact strongly with the surrounding anions, the exact position of the band depends upon the host crystal. Thus, the colors of Ce^{3+} and Eu^{2+} compounds, unusually for lanthanoid ions, vary with host structure.

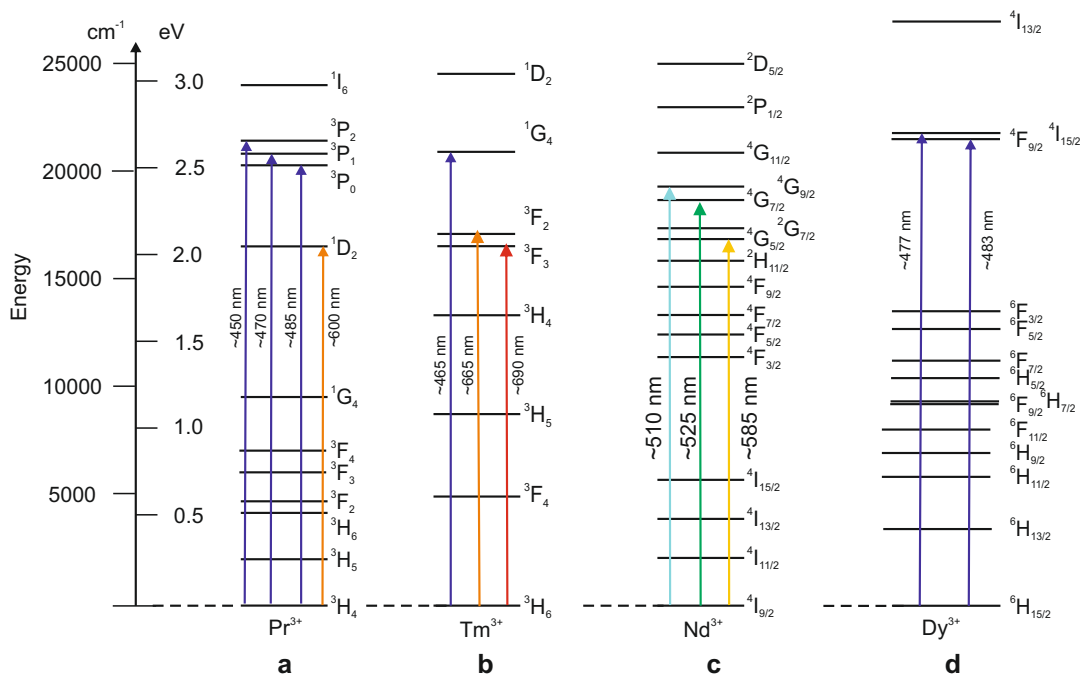
Pr^{3+} , Tm^{3+} , Nd^{3+} , and Dy^{3+}

The presence of Pr^{3+} ions generally colors the host matrix green. The principal transitions that contribute to the absorption spectrum are from the ground state 3H_4 to 3P_0 , at approximately

L

Lanthanoid Ion Color,
Fig. 1 Energy level diagrams (schematic) for (a) Ce^{3+} and (b) Eu^{2+} ions





Lanthanoid Ion Color, Fig. 2 Energy level diagrams (schematic) for (a) Pr^{3+} , (b) Tm^{3+} , (c) Nd^{3+} , and (d) Dy^{3+}

485 nm (blue-green), to $^3\text{P}_1$ at approximately 470 nm (blue) and to $^3\text{P}_2$ at approximately 450 nm (blue). A weak spin-forbidden transition to $^1\text{D}_2$ at approximately 590 nm (orange) is also present. Between these absorption peaks is a window of low absorption in the green region of the spectrum (Fig. 2a). The other “green” ion is Tm^{3+} . Here the main transitions are from the ground state $^3\text{H}_6$ to $^3\text{F}_3$ at approximately 690 nm (red), to $^3\text{F}_2$ at approximately 665 nm (red), and a weak spin-forbidden transition to $^1\text{G}_4$ at approximately 465 nm (blue) (Fig. 2b). Between these absorption bands is a green transmission window similar to that for Pr^{3+} . The Nd^{3+} ion tends to impart a lilac hue to materials. The main transition here is from the ground state $^4\text{I}_{9/2}$ to $^4\text{G}_{5/2}$ level, which absorbs strongly at approximately 585 nm (yellow), together with a transition to $^4\text{G}_{7/2}$ at approximately 525 nm (green) and $^2\text{G}_{9/2}$ at 510 nm (blue-green) (Fig. 2c). These absorption peaks remove the middle part of the optical spectrum, leaving both the violet and red extremes, so imparting a lilac hue to compounds. In the case

of Dy^{3+} with an electron configuration $4f^9$, there are two major absorption transitions, from the ground state $^6\text{H}_{15/2}$ to $^4\text{F}_{9/2}$ absorbing at 483 nm and to $^4\text{I}_{15/2}$, absorbing at 477 nm (Fig. 2d). Both of these remove blue from the spectrum, leaving materials containing Dy^{3+} to show a yellow coloration.

Some Applications of Lanthanoid Fluorescence Colors

Trichromatic Fluorescent Lamps

Trichromatic (color 80) fluorescent lamps use phosphors with active lanthanoid ions. Generally the lanthanoid ions can absorb a wide range of ultraviolet radiation efficiently, exciting the ions from the 4f-derived ground state to a broad band of energies formed from the interaction of 5d and 6s bands. Energy is then lost internally, in effect causing the phosphor matrix to warm slightly, until the sharp 4f-derived energy levels are reached. Photon emission between these energy

levels then occurs, giving color output. The favored red emitter is Eu^{3+} doped into Y_2O_3 ($\text{Y}_2\text{O}_3:\text{Eu}$), with the Eu^{3+} ions occupying the Y^{3+} sites. The ground state of Eu^{3+} is ${}^7\text{F}_0$. A transition from this state to the higher energy band, 5d band, absorbs the ultraviolet radiation given off by excited mercury atoms at 254 nm. Subsequent internal energy loss leaves the ion in the ${}^5\text{D}_0$ level. The main optical transition is between this level and ${}^7\text{F}_2$, producing emission at 611 nm (Fig. 3a). The green emission is from Tb^{3+} in host matrices $\text{La}(\text{Ce})\text{PO}_4$, $\text{LaMg}(\text{Ce})\text{Al}_{11}\text{O}_{19}$, or $\text{La}(\text{Ce})\text{MgB}_5\text{O}_{10}$, in which the Tb^{3+} ions replace La^{3+} . Tb^{3+} absorbs the mercury emission poorly, so it is coupled with a sensitizer, usually Ce^{3+} , which is able to absorb the 254-nm-wavelength mercury radiation efficiently by way of the 5d band. This absorbed energy is then transferred to the Tb^{3+} ions. The green emission, at a wavelength close to 540 nm, is mainly from a ${}^5\text{D}_4 \rightarrow {}^7\text{F}_5$ transition (Fig. 3b). Three other peaks of lesser intensity occur: ${}^5\text{D}_4 \rightarrow {}^7\text{F}_6$, 489 nm; ${}^5\text{D}_4 \rightarrow {}^7\text{F}_4$, 589 nm; ${}^5\text{D}_4 \rightarrow {}^7\text{F}_3$, 623 nm. The blue emission is produced by Eu^{2+} ions. The excitation and emission is directly to and from the 5d-derived band. The position of this band is strongly influenced by the host structure, and the usual tricolor lamp phosphor, $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}$, is chosen so as to have a suitable blue emission, with a maximum at 450 nm (Fig. 3c).

Plasma Displays

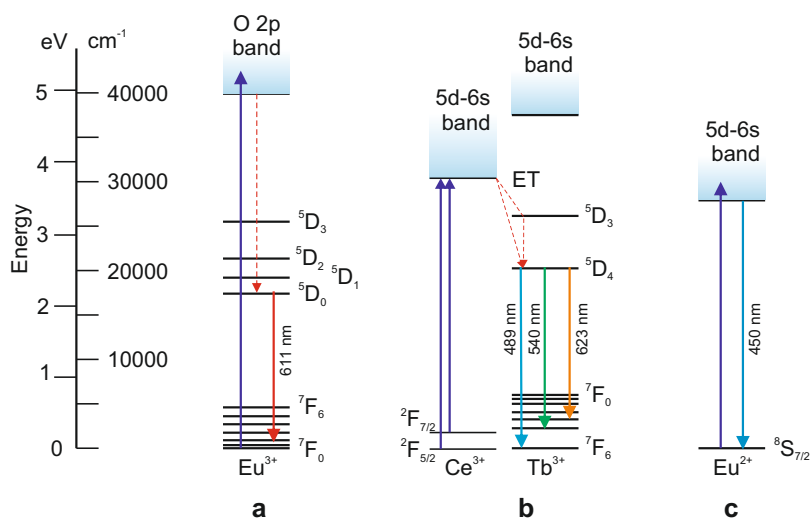
Plasma display panels are made up of a pair of glass plates containing a series of cells each of which acts as a miniature fluorescent lamp as described above. Each lamp is several hundred microns in size, and there are several million such lamps in a display. Each pixel consists of three lamps, giving off red, blue, and green lights. The working gas in the cells is a mixture of helium and xenon. When a high voltage is applied across two electrodes above and below a well, the gas is excited into a state that emits ultraviolet radiation, with principal wavelengths of 147 nm and 172 nm. Each well is coated internally with a red, green, or blue phosphor. The main lanthanoid phosphors used at present are yttrium gadolinium borate doped with europium ($\text{Y,Gd}\text{BO}_3:\text{Eu}^{3+}$, which gives a red emission, and barium magnesium aluminate doped with europium, $\text{BaMgAl}_{14}\text{O}_{23}:\text{Eu}^{2+}$, for blue emission (see Fig. 3). The green emission utilizes Mn^{2+} rather than a lanthanoid ion.

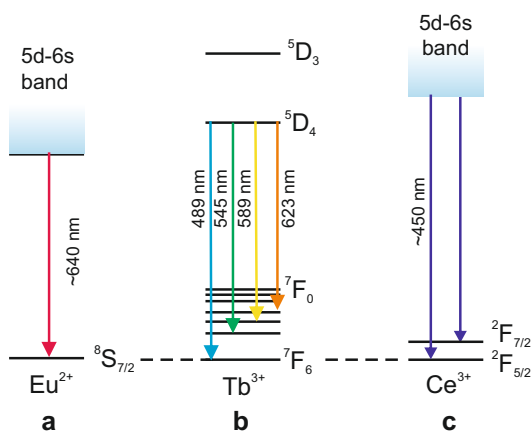
Phosphor Electroluminescent Displays

Electroluminescent displays containing a thin film of a phosphor, called thin-film electroluminescent (TFEL) displays, find use as flat panel color displays and backlighting in products such as instrument panels. The most promising devices use *ac* supplies in a thin-film electroluminescent (ACTFEL) display. Under the influence of an applied electric field, electrons enter the phosphor

Lanthanoid Ion Color,

Fig. 3 Energy level diagrams (schematic) for (a) Eu^{3+} , (b) $\text{Ce}^{3+}/\text{Tb}^{3+}$, and (c) Eu^{2+} in trichromatic lamp phosphors. ET energy transfer





Lanthanoid Ion Color, Fig. 4 Energy level diagrams (schematic) for (a) Eu^{2+} , (b) Tb^{3+} , and (c) Ce^{3+} in AC-powered thin-film electroluminescent (ACTFEL) displays

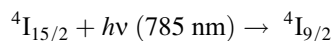
at the junction with a surface insulating coating. These are accelerated under the influence of the field until they collide with the luminescent centers in the phosphor, transferring energy in the process. The excited luminescent centers then fall back to the ground state and release energy by light emission.

Red emission is from calcium sulfide doped with europium ($\text{CaS}:\text{Eu}^{2+}$); the color is being generated by the transition from the 5d band to the ground state $^8\text{S}_{7/2}$ (Fig. 4a). At first sight this is surprising as the Eu^{2+} -derived tricolor lamp phosphor has a blue output. However, the position of the upper energy band depends upon the interaction of the d orbitals with the surrounding crystal, and in ZnS the softer bonding gives a broad emission centered close to 640 nm. Green emission is produced by zinc sulfide doped with terbium ($\text{ZnS}:\text{Tb}^{2+}$), with an output at a wavelength close to 545 nm, mainly from a $^5\text{D}_4 \rightarrow ^7\text{F}_5$ transition (Fig. 4b). Three other peaks of lesser intensity occur: $^5\text{D}_4 \rightarrow ^7\text{F}_6$, 489 nm; $^5\text{D}_4 \rightarrow ^7\text{F}_4$, 589 nm; $^5\text{D}_4 \rightarrow ^7\text{F}_3$, 623 nm. Blue emission still poses a problem for these displays, but the thiogallates CaSr_2S_4 , SrGa_2S_4 , and BaGa_2S_4 doped with the $4f^1$ ion Ce^{3+} are currently favored. The transitions between the 5d band and the $4f^1$ ground state doublet $^2\text{F}_{5/2}$ and $^2\text{F}_{7/2}$ are both in the blue region of the spectrum centered at 459 nm for the Ca

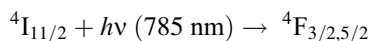
compound and 445 nm for the Sr and Ba phases (Fig. 4c).

Upconversion

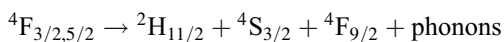
The conversion of infrared radiation to visible is variously known as upconversion, frequency upconversion, anti-Stokes fluorescence or cooperative luminescence. The majority of studies of upconversion have involved the lanthanoid ions Er^{3+} , Tm^{3+} , Ho^{3+} , and Yb^{3+} . At present the most efficient upconversion materials for infrared to visible conversion are lanthanoid fluorides such as NaLnF_4 doped with $\text{Yb}^{3+}/\text{Er}^{3+}$ or $\text{Yb}^{3+}/\text{Tm}^{3+}$ couples. The energy for upconversion can be gained by the active ion via several competing energy transfer processes. In principle, the simplest is for the active ion to pick up photons in two distinct steps. The first photon excites the ion from the ground state to an excited energy level, a process referred to as *ground state absorption* (GSA). A subsequent photon is then absorbed to further promote the excited ion to a higher energy level again, a process referred to as *excited state absorption* (ESA). The oxide CeO_2 doped with approximately 1% Er^{3+} exhibits upconversion in this way. Irradiation with near-infrared photons with a wavelength close to 785 nm excites the Er^{3+} ions from the $^4\text{I}_{15/2}$ ground state to the $^4\text{I}_{9/2}$ level:



Subsequent internal energy loss drops the energy to the $^4\text{I}_{11/2}$ and $^4\text{I}_{13/2}$ levels (Fig. 5a). The ions are then excited by ESA. Those in the $^4\text{I}_{11/2}$ energy level are excited to the $^4\text{F}_{3/2, 5/2}$ doublet:



These states subsequently relax to the $^2\text{H}_{11/2}$, $^4\text{S}_{3/2}$, and $^4\text{F}_{9/2}$ levels (Fig. 5b):

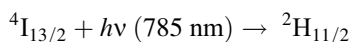
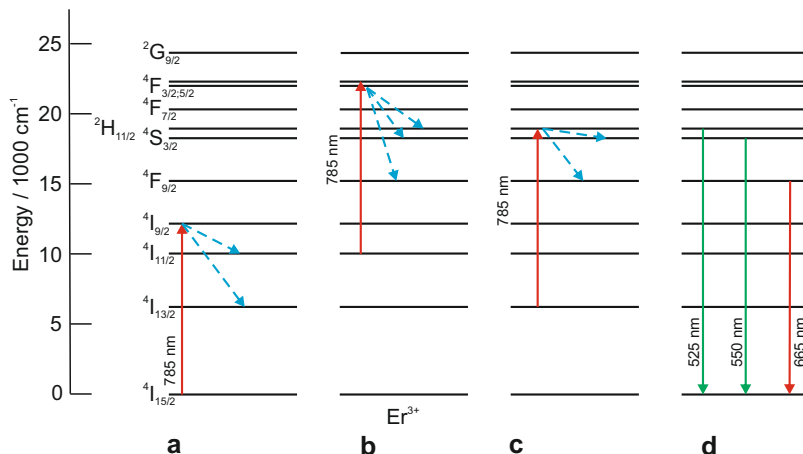


The ions in the $^4\text{I}_{13/2}$ energy level follow a similar path, being excited to the $^2\text{H}_{11/2}$ energy level:

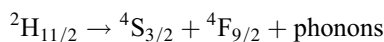
Lanthanoid Ion Color,

Fig. 5 Upconversion using Er^{3+} doped into CeO_2 , schematic:

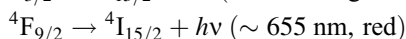
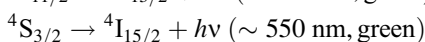
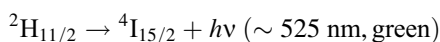
- (a) GSA + relaxation,
(b) ESA + relaxation,
(c) ESA + relaxation, and
(d) light emission



and then subsequently relax to the $^4\text{S}_{3/2}$ and $^4\text{F}_{9/2}$ levels (Fig. 5c):

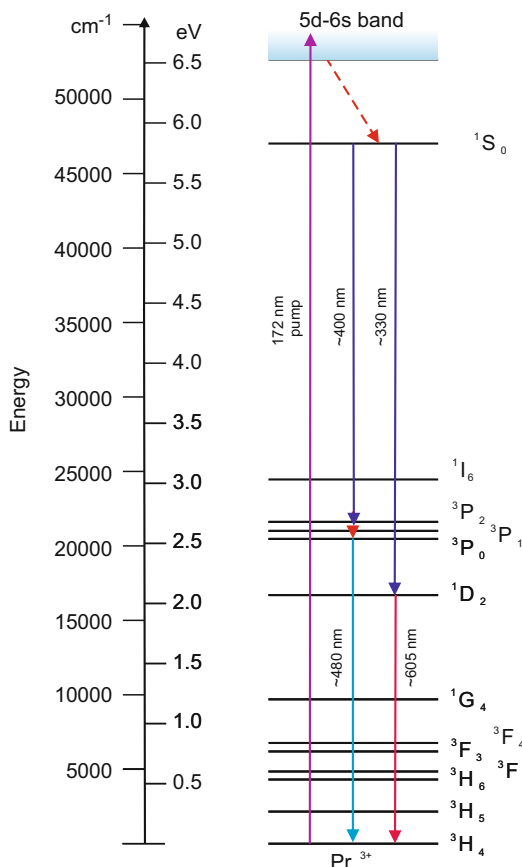


The result of this is that the levels $^2\text{H}_{11/2}$, $^4\text{S}_{3/2}$, and $^4\text{F}_{9/2}$ are populated to varying degrees, depending upon the precise details of the excitation and relaxation steps. Subsequent loss of energy from these levels gives rise to green and red emissions (Fig. 5d):



All upconversion spectra from Er^{3+} (including those using different mechanisms) are similar, but the relative intensities of the three peaks vary with concentration of the ions and the nature of the host matrix.

Energy transfer, in which input radiation is picked up by a sensitizer and is then passed to the emitter, is the mechanism of operation of host structures containing the co-dopants $\text{Yb}^{3+}/\text{Er}^{3+}$, which give a strong green emission, and $\text{Yb}^{3+}/\text{Tm}^{3+}$ which gives a blue emission. The ion that absorbs the incoming infrared radiation is the Yb^{3+} ion, which then transfers energy to Er^{3+} or



Lanthanoid Ion Color, Fig. 6 Visible light emission (quantum cutting) via photon cascade emission from Pr^{3+} ions irradiated by ultraviolet light

Tm³⁺ centers. Other upconversion mechanisms are also known.

Quantum Cutting

Quantum cutting is the reverse of upconversion, as one high-energy photon is processed (i.e., cut) to give out several lower-energy photons, typically transforming ultraviolet to visible. There are several mechanisms for quantum cutting; here photon cascade emission, exhibited by Pr³⁺ ions, is described. Initial absorption of high-energy 185-nm ultraviolet photons results in excitation to the 5d–6s energy band (Fig. 6). Subsequent relaxation takes the ion to the ¹S₀ level. Thereafter, the transitions giving rise to visible output are:

¹S₀ → ³P₃ at ~ 400 nm,
then ³P₀ → ³H₄ ground state at ~ 480 nm

¹S₀ → ¹D₂ at ~ 330 nm,
then ¹D₂ → ³H₄ ground state at ~ 605 nm

Other ions such as Tb³⁺ are also candidates for quantum cutting devices, but the mechanisms involved are more complex than that with Pr³⁺.

Cross-References

► [Transition Metal Ion Colors](#)

References

1. Tilley, R.J.D.: Chapter 7. In: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
2. Nassau, K.: Chapter 4. In: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
3. Huang, C.-H. (ed.): *Rare Earth Coordination Chemistry*. Wiley, Singapore (2010)
4. Linganna, K., Jayasankar, C.K.: *Luminescence Spectroscopy of the Lanthanides*. Scholars Press, Saarbrücken (2013)
5. Hänenen, P., Härmä, H. (eds.): *Lanthanide Luminescence*. Springer, Heidelberg (2011)
6. Cotton, S.: *Lanthanide and Actinide Chemistry*. Wiley, Chichester (2006)

Layering

► [Compositing and Chroma Keying](#)

Le Blon, Jacob Christoph

Rolf G. Kuehni
Charlotte, NC, USA



J. C. Le Blon, Head of a woman, ca. 1720 (three-color printing process)

Biography

Le Blon was born on May 2, 1667 in Frankfurt am Main, Germany, a descendant of Huguenots fleeing France in 1576, having settled there. His grandmother was a daughter of the artist and engraver Matthaus Merian the Elder (1593–1650). Showing an early interest in engraving and painting, he had, sometime between 1696 and 1702, an extended stay in Rome where he is reported to have studied art under the painter and engraver Carlo Maratta

(1625–1713) [1]. Around 1702 Le Blon moved to Amsterdam, where he worked as a miniature painter and engraver. In 1708/1709 he is known to have made colorant mixing experiments in Amsterdam, and in 1710 he made his first color prints with yellow, red, and blue printing plates. In 1717 he moved to London where he received a royal patent for the three-color printing and a related textile weaving process [2]. In 1722 he published a small book on painting, *Coloritto*, in French and English [3]. There he stated that “Painting can represent all visible objects with three colors, yellow, red, and blue.” During his stay in England, he produced several dozen images printed from three or four plates (the fourth for black ink) in multiple copies that initially sold well in England and on the continent. In the long run his enterprise did not succeed, however. Le Blon left England in 1735, moving to Paris where he continued producing prints by his method. In 1740 he began work on a collection of anatomical prints for which he had a solid list of subscribers. He died on May 16, 1741 in Paris. A detailed technical description of Le Blon’s method was published in 1756 by Antoine Gautier de Montdorge who supported him during his final years in Paris [4].

Major Accomplishments/Contributions

The idea of three chromatic primaries, yellow, red, and blue, was quite well established in Le Blon’s time among painters, graphically represented by Aguilonius in 1613 [5] and described by R. Boyle in 1664 (p. 220) [6]. What was new in Le Blon’s work is that he applied this concept to color printing of images in an entirely new fashion making greater and much subtler detailing and coloration possible. It required experience in deconstructing an image in terms of color so that printing multiple copies, based on only three or four plates, produced good quality coloration. It required the ability to mentally resolve the image into its presumed primary chromatic components and understanding and predicting the effects of superimposed printing inks in certain areas, for which extensive trial and error work was required. Le Blon manually engraved copperplates, using the mezzotint process, with the relative components of the three primary colors printed

successively in registration in the sequence blue, yellow, and red onto the paper substrate. As he gained experience, he at times used a fourth plate printing in black to achieve greater tonality and contrast, thus employing an early version of the CMYK process. Le Blon used the pigments Prussian blue, Stil de grain (yellow lake), a mixture of red lake and carmine for red, and a common printer’s black ink [4]. The pigments were dispersed in copal tree resin dissolved in copal oil to make the inks. The technical problems associated with the process prevented it from becoming a standard method and lithographic printing of color images from up to a dozen wood engravings or stones per image continued until H. E. Ives’ invention of the chromatic halftone printing process ca. 1890.

References

1. Lilien, O.M.: Jacob Christoph Le Blon, 1667–1741: Inventor of Three- and Four Colour Printing. Hiersemann, Stuttgart (1985)
2. Lowengard, S.: Jacob Christoph Le Blon’s system of three-color printing and weaving. In: *The Creation of Color in the 18th century Europe*. Columbia University Press, New York (2006). http://www.gutenberg-e.org/lowengard/C_Chap14.html. Accessed 11 May 2015
3. Le Blon, J. C.: *Coloritto, or the Harmony of Colouring in Painting: Reduced to Mechanical Practice* (with parallel French text). London (ca.1725)
4. Gautier de Montdorge, A.: *L’art d’imprimer les tableaux. Traité d’après les écrits, les opérations et les instructions verbales de J.-C. Le Blon*. Mercier, Paris (1756)
5. Aguilonius, F.: *Opticorum Libri Sex*. Plantin, Antwerp (1613)
6. Boyle, R.: *Experiments and Considerations Touching Colours*. Herringman, London (1664)

LED Sphere

- [Light Stage](#)

Light Bounces

- [Global Illumination](#)

Light Depreciation

► Lumen Depreciation

Light Distribution

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

Luminous intensity distribution

Definition

A luminaire property that indicates the values of the luminous intensities radiated in all relevant directions by the luminaire. Usually, the luminous intensities are given as intensities per 1000 lamp lumen when the lamp or lamps are operated under standard test conditions: 1/1000 lm.

Luminaire Performance Characteristic

The light distribution of a luminaire represents the most important performance characteristic of a luminaire. It is the basis for the determination of photometric characteristics of luminaires such as upward and downward light output ratio, utilization factors for defined areas or zones, beam angles, and glare figures. Luminaire manufacturers produce for their luminaires photometric datasheets which show these photometric characteristics together with a graphical representation of the light distribution itself.

Measurement

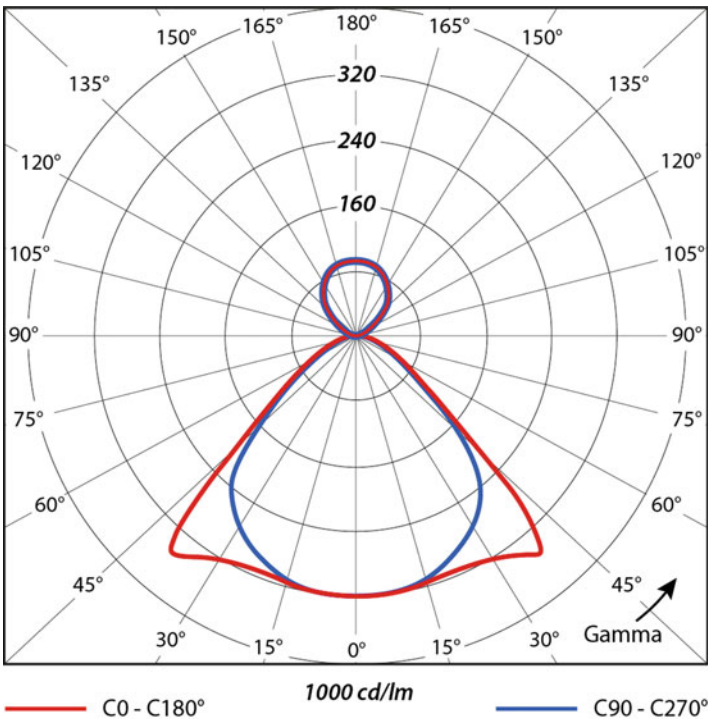
Strictly speaking, the light distribution can only be measured for a point source. In practice, the

measurement inaccuracy will be negligible provided the optical path length of the measuring setup is at least ten times the length of the light-emitting surface of the luminaire. For narrow beam projector type of luminaires, greater optical path lengths are required. The instrument used for measuring light distributions is called a goniophotometer. It exists of a mounting support for the luminaire, a photocell, and a, usually complex, system that rotates the combination of photocell and luminaire such that the intensities from all different light directions can be measured. Mirrors are used to keep the actual dimensions as small as possible while keeping the required optical path length. It is essential that the luminaire rotates in such a way that its normal operating position is at all times maintained. The measurements should be carried out under standard test conditions. These conditions refer to ambient temperature, air movement, mounting position, and lamp ballast type. The coordinating systems used for indicating the directions are either the so-called B-beta (usually used for projector type of luminaires) or the C-gamma system. The B angles of the B-beta system can be understood as the angles the pages of a book make when the book axis is held horizontally. The beta angles are angles in each page measured from the middle of the book axis. The C-gamma system in the same analogy can be understood as a book with its axis held vertically. C angles are the angles the pages make, and gamma angles are angles in the pages from the middle of the book axis.

Graphical Presentation

For the selection of suitable luminaires for a particular application, clear insight into the photometric quality of a luminaire is needed. This is why luminaire manufacturers produce photometric datasheets containing at least the polar intensity diagram. The polar luminous intensity diagram provides a good indication of the suitability of the light distribution for a certain application. In such diagram, the luminous intensities are presented using the C- γ system of coordinates in the form of curves in terms of cd/1000 lm of the

Light Distribution,
Fig. 1 Light distribution in the form of the polar luminous intensity diagram of the luminaire for which the I-table is given in Table 1



nominal lamp flux of the lamps employed (Fig. 1). Each curve represents one C plane. Curves are given for only the more important planes. For indoor lighting luminaires that are at least the planes through the two main axes of the luminaire. Rotationally symmetrical luminaires have the same light distribution in all C planes. The example of Fig. 1 shows a polar luminous intensity diagram of an indoor lighting luminaire that radiates its light in both the downward and upward direction. In the downward direction, it has a directional component in the plane across the longitudinal luminaire axis (C0–180° plane; red curve) and a rather diffuse component in the direction parallel to the longitudinal luminaire axis (C90–270°; blue curve). The upward component is diffuse in all planes which is the reason that the red and blue curves coincide.

Intensity Table

All lighting calculations of lighting installations are based on the light distribution of the

luminaires used. For this purpose the light distribution is given as a two-dimensional table, the so-called I-table. There are some different formats for these I-tables such as CIE, IESNA, Eulumdat, CIBSE, and some luminaire manufacturer’s specific formats. Eulumdat is the most popular format in Europe, whereas the IESNA format is more popular in the USA. The CIE format is an attempt to define a “recommended file format for electronic transfer of luminaire photometric data” on a global level, in order to overcome the regionally different formats currently in use. Many light calculation programs can use some different formats as input. Moreover, free software for conversion from one format to the other is available as well. The layout of the I-table is different for different types of luminaires. So have road lighting luminaire I-tables more values around the direction of the horizon than for other directions because these values very much determine glare, general lighting interior luminaire I-tables have a somewhat more regular layout because of the more diffuse character of these luminaires, and flood and accent lighting luminaires (projector type of luminaires)

Light Distribution, Table 1 I-table with the luminous intensity values for a specific luminaire (values are in cd/1000 lm)

Gamma	C0°	C15°	C30°	C45°	C60°	C75°	C90°
0°	322	322	322	322	322	322	322
10°	319	319	319	319	319	319	319
20°	313	312	311	309	306	304	302
30°	320	317	308	297	287	281	279
40°	352	345	326	301	269	245	234
50°	140	145	154	139	137	126	123
60°	66	66	66	68	65	63	62
70°	38	35	35	36	35	36	39
80°	19	20	22	22	22	19	19
90°	2.2	2.2	2.1	1.8	1.5	1.1	0.9
100°	6.1	6.1	6.2	6.2	6.1	5.9	5.8
110°	12	12	12	12	13	13	13
120°	22	22	22	23	23	24	24
130°	40	40	40	40	41	41	42
140°	60	60	60	60	60	61	61
150°	76	75	76	75	75	75	76
160°	84	84	84	84	84	85	85
170°	89	89	89	89	90	90	90
180°	91	91	91	91	91	91	91

have I-tables with extra values around the direction of the maximum beam intensity.

Table 1 shows part of an I-table in the C- γ coordinate system. The intensity tables usually give relative intensity values, viz., candela/1000 lamp lumen values to make the values solely dependent on the luminaire and not on the actual lumen of the lamp. For LED luminaires with non-replaceable LEDs, the intensity tables are usually given for the actual lumen that the luminaire emits. The luminaire used for the example of Table 1 is symmetrical relative to the C90–270° plane. For luminaires that do not have such symmetry, the intensity values for the planes C90 to C180° must also be given. Intensity tables used as input for calculation software have to be specified with smaller steps between the gamma and C angles.

Cross-Reference

► [Luminaires](#)

Light Pollution

Terry McGowan
Lighting Ideas, Inc, Cleveland Heights, OH, USA

Synonyms

[Glare](#); [Sky glow](#); Unwanted light especially referring to unwanted electric lighting at night

Definition

Outdoor electric lighting at night which is unwanted, unneeded, or wasteful and which results in glare, light trespass and sky glow or other harmful effects on people and the environment.

Light Pollution Recognized

Electric lighting makes travel, working, playing, and other normal living activities possible at

night. Indeed, the view of a lighted city from an airplane window or tall building is a hallmark of twenty-first-century civilization, yet little more than a century and a half ago, there were no such views and human activities outdoors at night were limited to those considered essential. Even then, they were carried out only with difficulty especially tasks involving work or travel. Limitations remained as candles and then gas flames and mantles began to generate more powerful and reliable night lighting, but the benefits were not widespread because of the relatively high cost of providing light especially over large outdoor areas at night.

When electric lighting was invented, one of its first uses was for outdoor illumination. Carbon arc lamps were used initially and then incandescent light sources, which were developed during the 1870s. These became the worldwide standard for all types of lighting until the commercial introduction of the fluorescent lamp in the late 1930s. Outdoor lighting, however, did not dramatically change until mercury lamps, the first of the so-called high-intensity discharge (HID) lamps became widely used for street, parking, and area lighting during the 1960s and then other types of HID lamps – high-pressure sodium (HPS) and metal halide (MH), due to their relatively higher efficacies, transformed the cost and use of outdoor lighting leading to rapid growth and the night time appearance of cities that we see today. Now, however, a new transformation is underway characterized by the growing use of solid state lighting in the form of Light Emitting Diodes (LEDs) which are increasingly being used for outdoor lighting.

Electric lighting has changed the night. And there are negative aspects that were recognized early on by critics who pointed out that bright lights frightened horses, annoyed pedestrians, and produced glare. There were those who also noticed that electric light changed the egg-laying habits of chickens, attracted insects, and modified the leaf and flowering characteristics of certain plants – early examples that hinted at more significant problems that later appeared due to unneeded or unwanted light. But the term “light pollution” did not appear until much later – in the 1980s. That’s when the large, powerful HID

lamps began to be used in large quantities for major outdoor lighting systems such as street lighting, parking areas, signs, sports stadiums, and floodlighting. The stray, uncontrolled light from these new sources as typically used outdoors plus the light from the headlights of cars and light escaping through windows from the interior lighting systems of buildings at night has now transformed the night environment in conspicuous ways including making the night sky less visible. Instead, light into the sky blankets cities with “sky glow,” a fog, or haze of light.

Astronomers, both professional and amateur, were among the first to recognize and talk about sky glow as light pollution as they tried to observe the stars from urban locations. As a first response, professional astronomers moved telescopes and observatories from cities to new locations in remote areas. Then, it was recognized that such places were increasingly rare and costly to develop. Further, most amateur astronomers were not able to move or take advantage of such sites. Members of the public who simply wanted to catch a glimpse of the Milky Way from their backyard or to show their children a comet or star constellation were not able to do so. The result is a growing and widespread reaction to uncontrolled electric lighting outdoors at night.

Efforts began, usually via zoning, planning, or the so-called “nuisance” regulations to limit and control unwanted light especially around astronomical observatories. It was found that light from urban areas could significantly affect the sky brightness 100–200 km away and that minimizing light directed into the sky was a simple and effective way to control such light. The term “light pollution” became familiar and efforts began to understand the various associated problems and to find ways to limit the detrimental effects. A global organization, the International Dark Sky Association (IDA), founded in 1988, focused on the task and continues to work “to preserve and protect the nighttime environment . . . through environmentally responsible outdoor lighting” [1]. Other regional and local organizations also appeared worldwide to raise awareness of the problems, write laws, provide educational materials, and organize local activities such as promoting the

use of shielded luminaires. Lighting technical and professional organizations have also recognized light pollution problems and have sponsored research, issued reports, produced recommended lighting practices, and written model legislation [2–4] in order to control it.

Types of Light Pollution

Light pollution was initially understood as primarily a visual problem caused by unnecessary or uncontrolled outdoor lighting that interfered with the viewing of the sky. But, there are now other considerations, and light pollution has been more broadly defined and classified as follows:

- **Light trespass** – Light generated in one location which, when poorly controlled, travels to another location where it is not wanted. An example would be a streetlight which is intended to light the street surface but also directs light into front yards or bedroom windows along the street.
- **Glare** – Excessive light directed into the eye where it causes visual discomfort or reduces the ability to see. Overly bright unshielded or misaimed luminaires are usually the main sources of glare.
- **Sky glow** – Light directed or reflected into the sky where it is scattered by air molecules, water, and other particles in the atmosphere and reflected diffusely to produce a fog or haze of brightness that reduces the visibility of the night sky.

Light pollution has energy implications since stray or misdirected light is luminous energy which has been converted from electric energy by the light source. That electric energy, the fuel, and other resources which produce it have therefore been wasted, delivering no value to those who pay for those resources.

The IDA estimates that some 30% of the electric energy used for outdoor lighting goes into the sky. Considering the hours of operation and an average electrical energy rate of \$0.10/kWh

(USA), the cost of that wasted energy is well over \$2 billion/year.

Spectral Effects

Electric lighting is the cause of light pollution. However, all light sources are not equally effective at generating light pollution of the types described. The scattering of light in the atmosphere that affects sky glow, for example, is a function of the wavelength of the light (Rayleigh scattering). Such scattering affects how far light pollution will travel. Similarly, the intensity (candlepower) of the luminaire as well as the way the light is aimed and directed matters too.

Astronomers who first sought to alleviate sky glow during the 1970s and 1980s found that the light from low-pressure sodium (LPS) lamps was helpful because all of the light from such lamps is emitted in a narrow range of wavelengths in the yellow part of the spectrum. A telescope equipped with a filter that absorbed those wavelengths and transmitted the rest of the spectrum would therefore see the sky as “dark.” US cities such as Tucson, AZ, and San Jose, CA, installed LPS street and area lighting as a way to maintain dark skies for area observatories while still providing essential lighting for drivers and pedestrians. LPS lamps, however, have poor color rendering and so there were objections to the light. Provisions were made for exceptions so that white light could be used in such applications as outdoor automobile sales lighting and sports facilities. Installations permitted as exceptions usually had other conditions attached that limited total lighting wattage, mandated certain operating hours, or required shielding – often in combinations. Of course, for the amateur astronomer or the casual viewer of the sky, outdoor lighting with LPS lamps and luminaires, which are typically unshielded, offers no advantage since the yellow light is visible to the eye and continues to produce light pollution via glare, light trespass, and sky glow (Fig. 1).

Note that while the spectral characteristics of the light source matter, overall light pollution, including sky glow, is the result of not only how



Light Pollution, Fig. 1 Example of the effect of light pollution on the visibility of the sky in a residential area near Toronto, Ontario, Canada. *Left:* during an area-wide power failure stars and constellations can be clearly seen.

Right: the same view after the power had been restored with streetlights, floodlights, and other outdoor lighting contributing to sky glow

L

the light is generated but also how it is modified by outdoor surfaces via reflection, refraction, and absorption. However, light pollution which interferes with a view of the sky in any given situation is always directly proportional to one quantity: total luminous power. Measured in lumens, luminous power is radiant power modified by the spectral sensitivity of the human eye. The key to controlling light pollution is to limit or reduce the total lumens being generated from light sources and sent into the night environment. Turning lights off when not needed, shielding luminaires, and reducing illumination levels to the minimum values required are all techniques that can be used to minimize light pollution problems. These are common-sense ideas, and there are many more [5, 6] although what should happen, of course, is that lighting should be designed to incorporate the long-established principles of proper lighting design which include the requirements to control stray light, minimize glare, and not use excessive light for the intended application. The problem is

that much outdoor lighting is not “designed” in the sense that major buildings are designed with careful thought given to materials, appearance, function, and purpose. Rather, outdoor lighting is often a make-do or do-it-yourself affair. It “happens” or factors of cost, expediency, or lack of information drive the installation (Fig. 2).

Other Light Pollution Effects

Research reported during the 1990s has shown that light pollution seriously affects the health and behavior of animals, including mammals, birds, amphibians, and insects. Changes in migration patterns, food gathering and feeding habits, mating, and social behavior have all been documented [7, 8]. However, bird and sea turtle problems have received the most attention.

Birds may become disorientated by lighting to the extent that they die from exhaustion while continuously circling lighted buildings or,

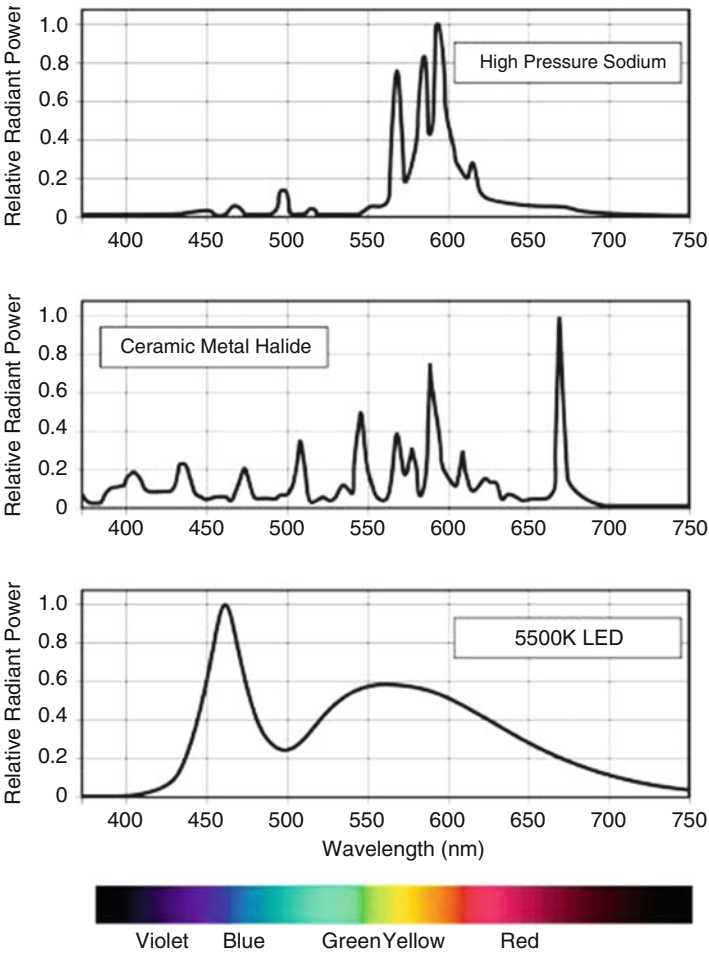
Light Pollution, Fig. 2

A comparison of the spectra of light sources commonly used for outdoor lighting.

Top: high-pressure sodium is the most widely used and is characterized by its familiar orange color.

Middle: metal halide lamps emit white light but can have a visually “warm” or “cool” color depending upon the chromaticity rating. Emissions, however, are relatively balanced over the visible spectrum.

Bottom: white-light LED light sources use blue-emitting LEDs to activate phosphors which emit the light shown in the 500 nm and above range. These so-called “blue-rich white-light” LEDs are of concern because light and human health research indicates that eye receptors which control circadian rhythms have a peak sensitivity near 460 nm, similar to the emission curve of the blue LEDs



confused, they collide with the buildings. According to a Chicago Audubon Society study [9] documented by actual counts of dead birds found at the base of tall buildings, bird deaths due to building collisions were reduced significantly when just the decorative floodlighting of buildings was switched off for most of the night hours during migration seasons.

Baby sea turtles hatch on beaches and have a short time to dig out of the sand and to find their way to the sea before they die by dehydration or are eaten by predators. The turtles appear to use the brightness of the horizon as a visual clue to find the water. Area and streetlighting along beachfront areas confuses the turtles, so efforts have been made to remove, turn off, or modify such systems including changing to amber-colored light sources which, because of their

spectral sensitivity, appear less bright to turtle eyes [10].

Human Health Effects

A controversial issue involving light pollution is whether or not outdoor lighting negatively affects human health. Research beginning in the 1970s has indicated that certain uses and kinds of electric light and, particularly, light at night (LAN) when the human body expects and needs darkness for sleep, can be detrimental to health because it interferes with the body’s normal circadian rhythms.

These rhythms mark the periods of alertness and sleep due to the ebb and flow of hormones, such as melatonin, in the body. According to the

research, strong circadian rhythms are essential for overall health and mental functioning and particularly for cell repair which helps protect against major diseases including cancer. The general rule is that people need bright days and dark nights so current research is exploring questions about light intensity, spectrum, timing, and exposure duration to determine the “dose” of light that might result in the interruption of the melatonin cycle and, further, what kind of interruptions lead to an increased risk of disease. The research literature on the subject is already extensive – one compilation now lists over 1,000 citations [11], and how outdoor lighting impacts human health will remain an important subject of research and discussion.

The American Medical Association adopted a resolution in 2009 entitled “Advocating and Support for Light Pollution Control Efforts and Glare Reduction for Both Public Safety and Energy Savings” [12]. Significantly that AMA view was broadened in 2012 when they said, “The (AMA) policy also supports the need for developing lighting technologies that minimize circadian disruption and encourages further research on the risks and benefits of occupational and environmental exposure to light at night” [13].

Outdoor lighting, because such lighting impacts the natural environment and so many people see it, use it, and experience it, has become complicated. It is now a social, technical, medical, and environmental problem that has to be carefully analyzed rather than a traditional exercise in illuminating engineering so that people can see. Light pollution issues contribute to outdoor lighting’s complexity, but light pollution also calls attention to important and growing outdoor lighting problems that have to be addressed and resolved.

References

1. Mission Statement. The International Dark Sky Association (IDA). www.darksky.org. Accessed 13 December 2014

2. Model Lighting Ordinance (MLO): A joint effort of the International Dark Sky Association and the Illuminating Engineering Society of North America. www.darksky.org or www.ies.org (2011). Accessed 13 December 2014
3. International Commission on Illumination: Technical Report – Guide on the Limitation of the Effects of Obtrusive Light from Outdoor Lighting Installations. CIE 150. CIE Publications. <http://www.cie.co.at/index.php/Publications> (2003)
4. International Commission on Illumination: Technical Report – Guidelines for Minimizing Sky Glow. CIE 126. CIE Publications. <http://www.cie.co.at/index.php/Publications> (1997)
5. Mizon, B.: *Light Pollution: Responses and Remedies*, 2nd edn. Springer, New York (2012) ISBN-13 978-1461438212
6. Rich, C., Longcore, T. (eds.): *Ecological Consequences of Artificial Night Lighting*. Island Press, Washington, DC (2005)
7. The Royal Commission on Environmental Pollution: *Artificial light in the environment*. The Stationery Office, United Kingdom (2009)
8. Chicago Audubon Society: *Lights out after 11:00 p.m. During migration*. <http://www.chicagoaudubon.org/lightsout.shtml>. Accessed 20 June 2012
9. National Geographic News: *Saving sea turtles with a lights out policy in Florida*. http://news.nationalgeographic.com/news/2003/03/0310_030310_turtlelight.html. Accessed 20 June 2012
10. International Dark Sky Association: *Visibility, environmental and astronomical issues associated with Blue-Rich White Outdoor Lighting*. Tucson, IDA (2010) <http://darksky.org/assets/documents/Reports/IDA-Blue-Rich-Light-White-Paper.pdf>
11. Wagner, R.: *Light at night, human health – references with abstracts*. http://www.trianglealumni.org/mcrol/References-With_Abstracts.pdf (2011). Accessed 12 Jun 2012
12. American Medical Association House of Delegates: *Resolution: 516 (A-09)* http://www.eficienciaenergetica.gub.uy/con_luminica/American%20Medical%20Association%20-%20Resolution%20516.pdf (2009). Accessed 10 June 2012
13. American Medical Association: *AMA Adopts New Policies at Annual Meeting*. <http://www.ama-assn.org/ama/pub/news/news/2012-06-19-ama-adopts-new-policies.page> (2012). Accessed 2 July 2012
14. Bogard, P.: *The End of Night: Searching for Natural Darkness in an Age of Artificial Light*, First edn. Little, Brown and Co, New York (2013)

Light Sensors

► Photodetector

Light Stage

Abhijeet Ghosh

Department of Computing, Imperial College
London, London, UK

Synonyms

LED sphere; Lighting arc; Lighting dome

Definition

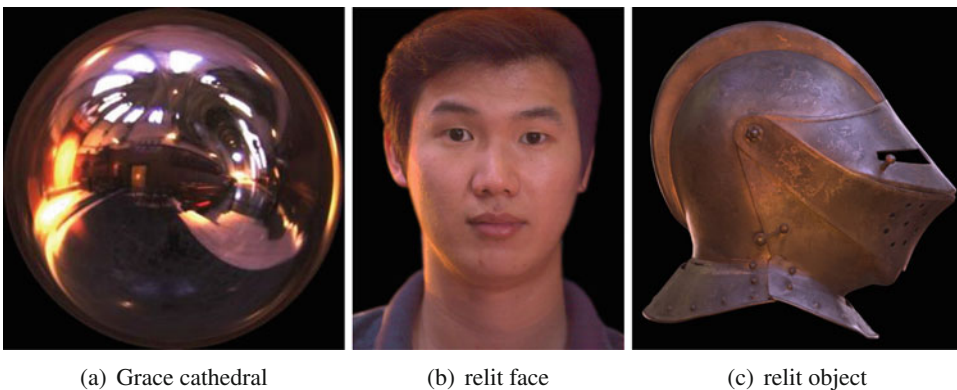
A Light Stage is a device to illuminate an object or a subject under controlled illumination from a full sphere of surrounding directions. Light Stage illumination control, with control over light intensity, color, or direction, is usually performed in conjunction with photographing an object or subject's reflectance response to incident illumination. This process of recording a subject under controlled illumination has various applications in computer graphics and vision such as relighting, lighting reproduction, or reflectometry.

Light Stage and Reflectance Field

The Light Stage was first introduced by Debevec et al. [2] for acquiring the reflectance field of a

human face. The process involved spiraling a spotlight source tracing a spherical path around a subject using a rope and pulley system while recording the subject's face with a video camera. The whole process required around a minute to complete light movement covering a full sphere of incident directions. This enabled the camera to record the response of the subject's face under a dense set of discrete lighting directions, with the set of acquired photographs referred to as a 4D reflectance field (2D for incident illumination direction $\times$ 2D for each photograph recorded by the camera). Such a 4D reflectance field enables image-based relighting of the acquired subject using a simple dot product of the reflectance field data with a desired lighting (represented as a spherical environment map) (see Fig. 1).

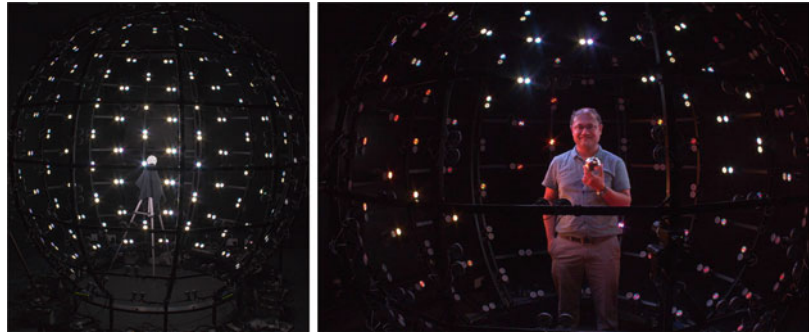
Various versions of the Light Stage have subsequently been developed for efficient capture of subject reflectance with applications in film visual effects [1]. Light Stage 2 was developed as a semicircular arc of strobe lights that could be rotated rapidly around a subject in 8 seconds. The device could be employed to rapidly acquiring facial reflectance fields of subjects in various facial expressions to build an animatable character [6]. This process was further extended with Light Stage 5 to record an actor's dynamic facial reflectance field using time-multiplexed illumination and high-speed photography [19]. This required a change in the Light Stage design, with a set of 156 bright white LEDs mounted on a 2 m



Light Stage, Fig. 1 Image based relighting of a reflectance field acquired using a Light Stage [16]. (a) Grace cathedral lighting environment. (b, c) Relighting results

for: (b) a face acquired in Light Stage 5 (156 images), and (c) a medieval helmet acquired in Light Stage 6 (253 images)

Light Stage, Fig. 2 LED sphere design for a Light Stage with LEDs mounted along latitudes-longitudes [9]. Lighting reproduction on a subject using RGB LEDs on the LED sphere [8]



(a) LED sphere

(b) Lighting reproduction

diameter geodesic structure surrounding the subject. Finally, reflectance field acquisition was further extended for full human body using Light Stage 6, a large 8 m diameter dome (two thirds of a full sphere) with more than 6000 bright white LEDs [4]. Figure 2a shows an alternate design of an LED sphere where, instead of a standard icosahedral subdivision of a sphere, 168 LED lamps have been mounted along latitudes-longitudes [9].

Lighting Reproduction

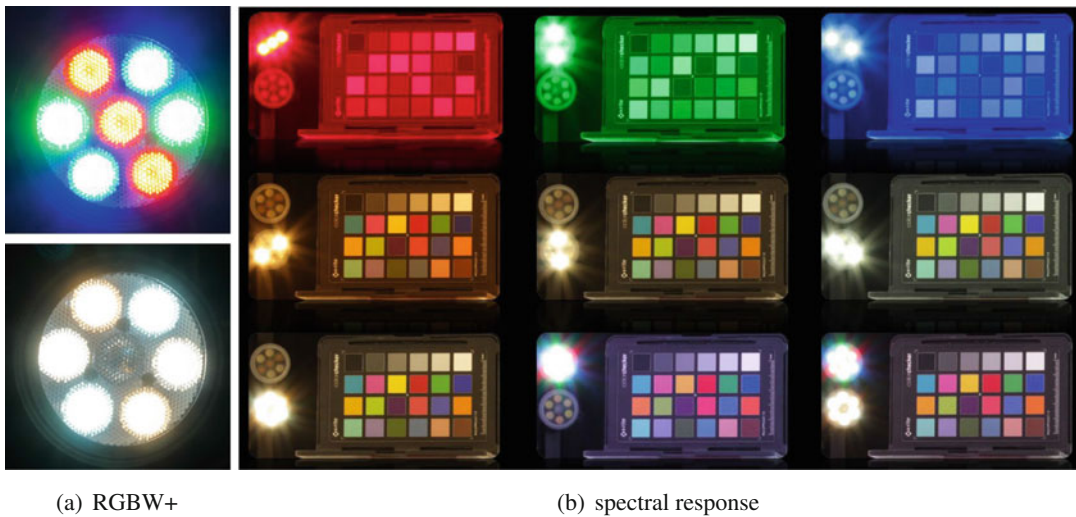
Besides image-based relighting, Light Stages can also be employed for direct lighting reproduction applications where an actor is lit inside a Light Stage with a desired lighting environment. This enables compositing the actor's imagery or performance into a virtual background with matching illumination. Debevec et al. [3] first proposed this process with Light Stage 3, which consisted of a geodesic spherical structure (same as Light Stage 5) with 156 programmable RGB LEDs. The device could be programmed to reproduce the colors and intensities of an HDR lighting environment on an actor at video rates. An example of this can be seen in Fig. 2b. The method has been employed in several film visual effects pipelines for illuminating actors in a studio environment for realistic digital compositing, e.g., in the film *Gravity* to illuminate actors wearing space suits with light from the surrounding space and Earth as seen from orbit [14].

More recently, the method of RGB illumination has been extended for improved spectral

lighting reproduction using a wider set of color LEDs including red (R), green (G), blue (B), cyan (C), amber (A), and white (W) by LeGendre et al. [11]. They demonstrated more accurate color reproduction with such an extended RGBCAW lighting compared to standard RGB lighting using USC-ICTs multispectral Light Stage X apparatus. A similar multispectral lighting capability has recently been achieved using a combination of RGB, and an extended set of white (W+) LEDs spanning warm, neutral, and cold color temperatures respectively [8] (see Fig. 3).

Facial Capture

More recently, Light Stage systems have been employed for high-quality facial capture including acquisition of detailed facial geometry and reflectance maps for realistic facial renderings. Such 3D facial geometry and reflectance data are better suited for standard visual effects rendering pipelines than the image-based data acquired for relighting or lighting reproduction applications. Debevec et al. first explored estimating 3D shape (in the form of a surface normal map) and diffuse and specular reflectance from the dense reflectance field of a face [2]. This process was further refined by Weyrich et al. [20] who employed an LED sphere with 150 lighting directions to acquire dense reflectance information of a face from 16 viewpoints for detailed facial reflectance modeling. However, these initial works required dense capture of facial reflectance data to compute



Light Stage, Fig. 3 Multispectral lighting using RGBW+ LEDs [8] (a), and their corresponding spectral response on an Xrite color chart (b)

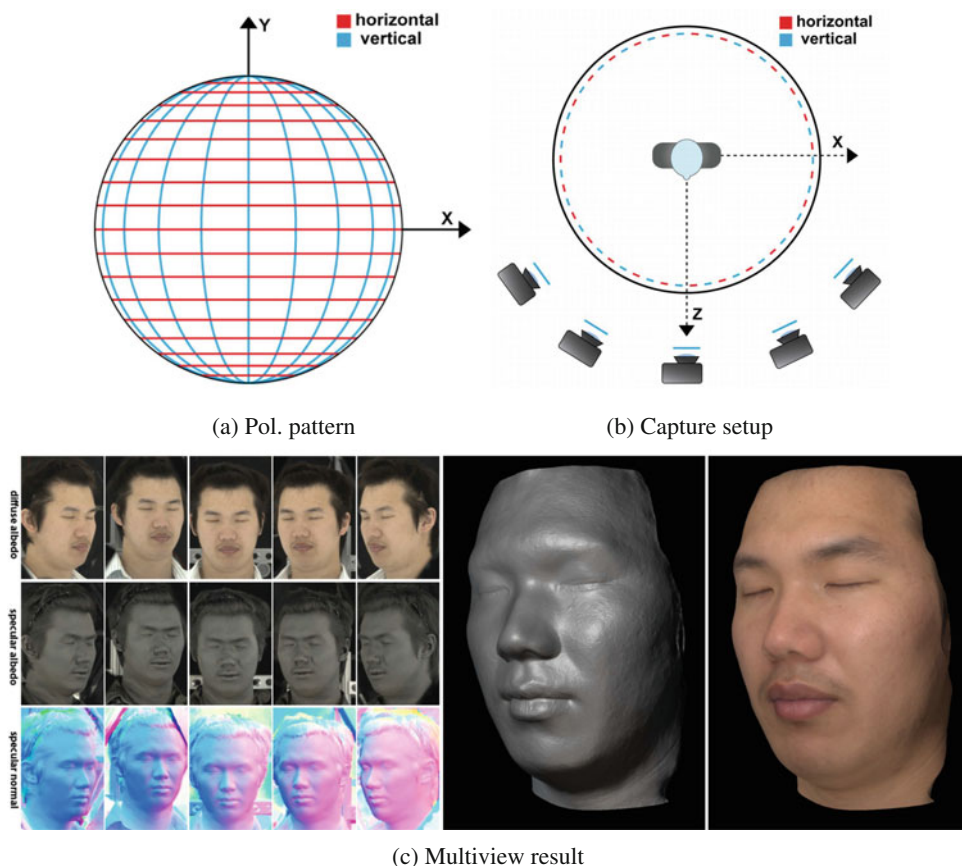
the reflectance and normal maps for rendering (Fig. 4).

Ma et al. [12] introduced a step change in facial reflectance acquisition process with a Light Stage. They employed the Light Stage 5 LED sphere to rapidly illuminate a subject with four spherical gradient illumination conditions (one full-on uniform illumination, and three principal linear gradients in the X, Y, and Z directions) to estimate high-quality facial reflectance and normal maps. They further proposed polarizing all the lights on the LED sphere in a specific pattern to allow high-quality separation of diffuse and specular reflectance and normal maps from a frontal viewpoint (in conjunction with mechanical flipping of a linear polarizer in front of the camera). Besides high-quality facial rendering with hybrid normal mapping, the high resolution surface details estimated from the specular normal map could then be employed for very high resolution facial geometry reconstruction (including fine-scale surface details, such as skin pore and fine wrinkles). The capture method was however restricted to a single frontal viewpoint due to the polarization pattern on the LED sphere. This restriction was lifted by Ghosh et al. [5], who proposed a multiview polarization solution for facial capture. Their solution

involves the LEDs on the Light Stage to be partitioned into two orthogonal polarization banks producing lines-of-latitude (horizontal) polarization and lines-of-longitude (vertical) linear polarization, respectively. This allows high-quality diffuse-specular separation from camera viewpoints (with a fixed vertical linear polarizer) placed all around the equator of the LED sphere, enabling high-quality multiview facial capture. While the method was originally demonstrated using interleaved banks of lat-long polarization with Light Stage 5, it was subsequently adapted for facial capture using Light Stage X with individually controllable polarization banks in each lighting unit.

Facial capture with polarized spherical gradient illumination has been employed extensively for film visual effects, starting with scanning in Light Stage 5 for AVATAR and subsequent multiview facial capture in Light Stage X for many more recent films. The technique and apparatus have been recently recognized by the Academy of Motion Picture Arts and Science with a Technical Achievement Award in 2019 [15].

Under the assumption of pure diffuse surface reflectance, binary spherical gradient illumination has instead been employed for estimating



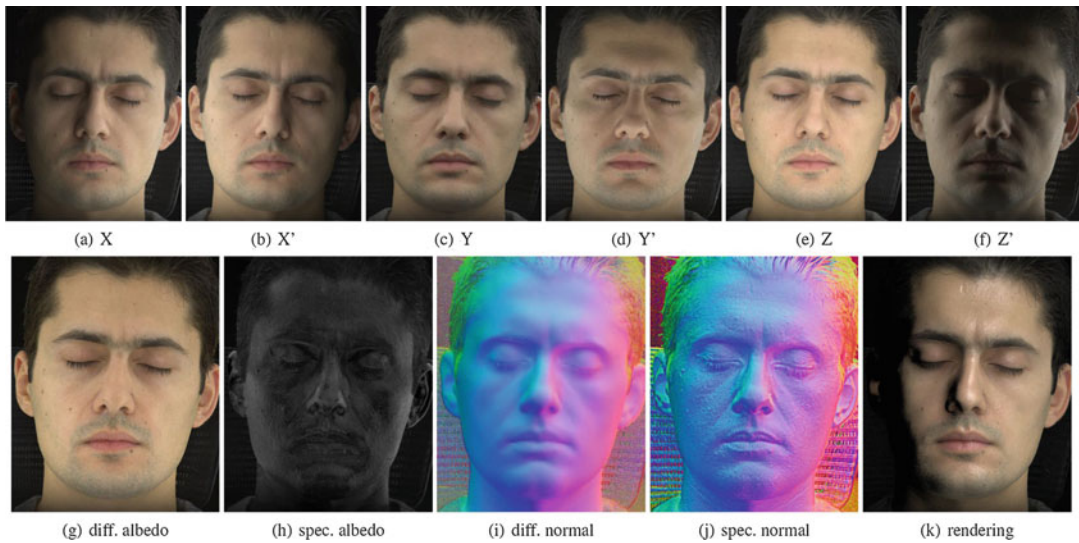
Light Stage, Fig. 4 Multiview facial capture with polarized spherical gradient illumination [5]. (a, b) LED sphere latitude-longitude polarization pattern and multiview

capture setup. (c) Acquired data with diffuse-specular separation, and the resulting reconstructed facial geometry and facial rendering

photometric normals for full-body performance capture (using Light Stage 6) [18]. Kampouris et al. [9] have recently demonstrated how binary spherical gradient illumination can also be employed for high-quality diffuse-specular separation of reflectance and photometric normals using color-space analysis (see Fig. 5). An advantage of this method over polarized spherical gradient illumination is that it reduces data capture by half, and the acquisition is more light-efficient and completely view-independent. Spectral multiplexing of the binary gradients further reduces data capture to just two photographs for view-independent diffuse-specular separation of albedo and photometric normals. Kampouris et al. employed a versatile multispectral Light Stage built with off-the-shelf components [8] for this purpose.

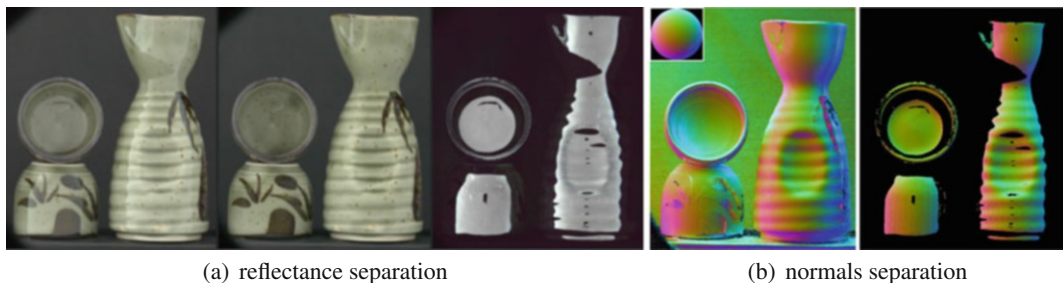
Object Capture

Specific Light Stage designs have also been developed for acquiring objects, particularly those exhibiting high frequency reflectance functions such as sharp specularity, or transparent objects casting sharp caustics. Hawkins et al. [7] introduced a dual Light Stage design that replaced a traditional Light Stage that usually has a discrete set of lights with a continuous spherical diffuse dome enclosing an object. The dual Light Stage exploits the property of reciprocity of light transport to indirectly record a reflectance function at very high resolution. It does this by placing an actuated laser source at the position of a desired virtual camera which illuminates the object with individual beams of laser light through a small hole in the dome. The object then reflects or



Light Stage, Fig. 5 A face acquired with binary spherical gradients and their complements (a–f) [9], and estimated diffuse and specular albedo (g, h) and photometric normal

maps (i, j), which enable realistic relighting with hybrid normal rendering (k)



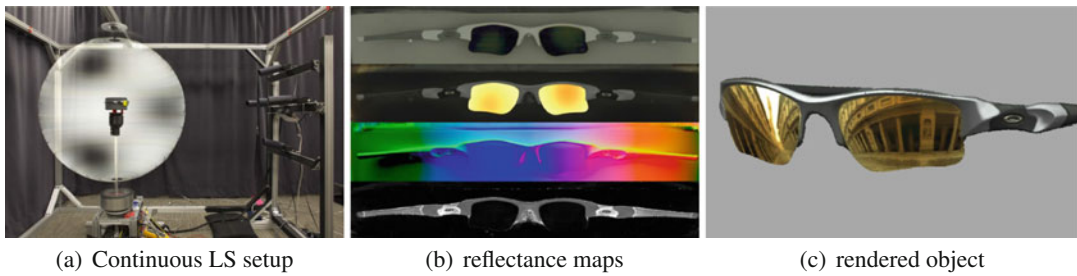
Light Stage, Fig. 6 Diffuse-specular separation of reflectance (a) and photometric normals (b) with structured environmental illumination using a reflective Light Stage [10]

scatters the beam onto the diffuse spherical dome from where the reflectance distribution is seen by a camera placed above that records the dome reflection using a fish-eye lens. Reciprocity is used afterward to convert the camera image into a continuous reflectance function while the corresponding laser beam location is converted into a virtual camera pixel. The method allows relighting of highly specular or transparent objects without any aliasing artifacts that would be visible when employing a traditional Light Stage.

Peers et al. [13] proposed a simpler design of a reflective Light Stage that employs a projector with a fish eye lens and a reflective hemispherical

dome to illuminate sharp specular objects with continuous illumination covering the front hemisphere. The setup could be employed for image-based relighting using reflectance field acquisition, as well as reflectance separation (into diffuse and specular components) using structured environmental illumination [10] (see Fig. 6).

Besides acquisition of highly specular objects for image-based relighting, a continuous illumination Light Stage [17] has also been designed for acquiring surface reflectance (with diffuse-specular separation) and shape (in the form of normal map and surface geometry) of such objects



Light Stage, Fig. 7 A continuous Light Stage for illuminating specular objects with spherical harmonic illumination [17]. (a) Physical setup. (b) Acquired reflectance and

normal maps of Oakely sunglasses. (c) 3D rendering of sunglasses with acquired data

for standard rendering pipelines. Tunwattanapong et al. proposed employing a rotating LED arc to cover a full sphere of surrounding directions with continuous spherical illumination. Furthermore, they employed high frequency spherical harmonic illumination to separate reflectance of specular objects into diffuse and specular components, as well as estimate surface normals and spatially varying specular roughness over the sample for realistic rendering. Combining the setup with multiview acquisition enabled them to also acquire 3D geometry of such objects for rendering from a novel viewpoint (Fig. 7).

References

1. Debevec, P.: The light stages and their applications to photoreal digital actors. In: SIGGRAPH Asia Technical Briefs (2012)
2. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH'00). ACM Press/Addison-Wesley Publishing Co., USA, 145–156. (2000). <https://doi.org/10.1145/344779.344855>
3. Debevec, P., Wenger, A., Tchou, C., Gardner, A., Waese, J., Hawkins, T.: A lighting reproduction approach to live-action compositing. ACM Trans. Graph. (Proc. SIGGRAPH). **21**(3), 547–556 (2002)
4. Einarsson, P., Chabert, C.F., Jones, A., Ma, W.C., Lamond, B., Hawkins, T., Bolas, M., Sylwan, S., Debevec, P.: Relighting human locomotion with flowed reflectance fields. In: Proceedings of the 17th Eurographics Conference on Rendering Techniques, EGSR'06, pp. 183–194. Eurographics Association, Goslar Germany (2006). <http://dl.acm.org/citation.cfm?id=2383894>. 2383917
5. Ghosh, A., Fyffe, G., Tunwattanapong, B., Busch, J., Yu, X., Debevec, P.: Multiview face capture using polarized spherical gradient illumination. ACM Trans. Graph. (TOG). **30**(6), 129 (2011)
6. Hawkins, T., Wenger, A., Tchou, C., Gardner, A., Göransson, F., Debevec, P.: Animatable facial reflectance fields. In: Proceedings of the Fifteenth Eurographics Conference on Rendering Techniques, EGSR'04, pp. 309–319. Eurographics Association, Aire-la-Ville (2004). <https://doi.org/10.2312/EGWR/EGSR04/309-319>
7. Hawkins, T., Einarsson, P., Debevec, P.: A dual light stage. In: Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques, EGSR '05, pp. 91–98 (2005). <https://doi.org/10.2312/EGWR/EGSR05/091-098>
8. Kampouris, C., Ghosh, A.: ICL multispectral light stage: Building a versatile LED sphere with off-the-shelf components. In: Proceedings of the Eurographics 2018 Workshop on Material Appearance Modeling, EG MAM '18, pp. 1–4 (2018). <https://doi.org/10.2312/mam.20181190>
9. Kampouris, C., Zafeiriou, S., Ghosh, A.: Diffuse-specular separation using binary spherical gradient illumination. In: Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations, SR '18, pp. 1–10 (2018). <https://doi.org/10.2312/sre.20181167>
10. Lamond, B., Peers, P., Ghosh, A., Debevec, P.: Image-based separation of diffuse and specular reflections using environmental structured illumination. In: Proceeding of International Conference on Computational Photography (2009)
11. LeGendre, C., Yu, X., Liu, D., Busch, J., Jones, A., Pattanaik, S., Debevec, P.: Practical multispectral lighting reproduction. ACM Trans. Graph. (Proc. SIGGRAPH). **35**(4), 32:1–32:11 (2016). <https://doi.org/10.1145/2897824.2925934>
12. Ma, W.C., Hawkins, T., Peers, P., Chabert, C.F., Weiss, M., Debevec, P.: Rapid acquisition of specular and

- diffuse normal maps from polarized spherical gradient illumination. In: Proceedings of the 18th Eurographics conference on Rendering Techniques (EGSR'07). Eurographics Association, Goslar, DEU, 183–194. (2007).
13. Peers, P., Hawkins, T., Debevec, P.: A reflective light stage. Tech. Rep. (2006)
 14. Seymour, M.: Gravity: vfx that's anything but down to earth (2013). <https://www.fxguide.com/fxfeatured/gravity/>
 15. Seymour, M.: SciTech awards: scanning Hollywood with Paul Debevec (2019). <https://www.fxguide.com/fxfeatured/scitech-awards-scanning-hollywood-with-paul-debevec/>
 16. Tunwattanapong, B., Ghosh, A., Debevec, P.: Practical image-based relighting and editing with spherical-harmonics and local lights. In: Proceedings of the 2011 Conference for Visual Media Production, CVMP'11, pp. 138–147 (2011). <https://doi.org/10.1109/CVMP.2011.22>
 17. Tunwattanapong, B., Fyffe, G., Graham, P., Busch, J., Yu, X., Ghosh, A., Debevec, P.: Acquiring reflectance and shape from continuous spherical harmonic illumination. ACM Trans. Graph. (Proc. SIGGRAPH). **32**(4), (2013)
 18. Vlastic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Trans. Graph. **28**(5), 174 (2009)
 19. Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., Debevec, P.: Performance relighting and reflectance transformation with time-multiplexed illumination. ACM Trans. Graph. (Proc. SIGGRAPH). **24**(3), 756–764 (2005)
 20. Weyrich, T., Matusik, W., Pfister, H., Bickel, B., Donner, C., Tu, C., McAndless, J., Lee, J., Ngan, A., Jensen, H.W., Gross, M.: Analysis of human faces using a measurement-based skin reflectance model. ACM Trans. Graph. **25**(3), 1013–1024 (2006)

Light, Electromagnetic Spectrum

Joanne Zwinkels

National Research Council Canada, Ottawa, ON, Canada

Synonyms

Electromagnetic spectrum – Optical spectrum; Light – Light stimulus, Optical radiation, Electromagnetic radiation

Definitions

Light is so much a part of our everyday existence contributing to our quality of life, as well as a key enabler in photonics, solar power, and new lighting technologies, that the year 2015 was declared by the UNESCO as the International Year of Light and Lighting Technologies.

Light has many different meanings depending upon the application.

When it is used to broadly describe *optical radiation*, it is defined as electromagnetic radiation with wavelengths between approximately 10 nm and 1 mm, i.e., a term that is used to describe the ultraviolet, visible, and infrared regions of the electromagnetic spectrum.

When it is used to describe a *light stimulus*, the International Lighting Vocabulary [1] gives the following two definitions: (1) It is a characteristic of all sensations and perceptions that is specific to vision and (2) it is radiation that is considered from the point of view of its ability to excite the human visual system, i.e., it is limited to electromagnetic radiation with wavelengths in the visible spectral region between approximately 380 and 780 nm.

The term *light* is also sometimes used as a synonym for *electromagnetic radiation* (emr). However, emr encompasses not only the ultraviolet, visible, and infrared range but also the X-ray and gamma-ray ranges <10 nm and the microwave and radiowave ranges >1 mm. The use of the term *light* for this extended emr range of wavelengths should be avoided.

Electromagnetic radiation is the emission or transfer of energy in the form of electromagnetic waves and in the form of photons.

Photon is a quantum of electromagnetic radiation that is usually associated with radiation that is characterized by one wavelength or frequency (monochromatic radiation).

Electromagnetic wave is a transverse oscillation of inextricably linked electric and magnetic fields traveling through space; through empty space, it travels at the speed of light in vacuum; through other media, its speed is reduced by the value of the refractive index of the medium.

Electromagnetic spectrum is defined as the graphical representation of electromagnetic waves arranged according to their wavelength.

Overview

Early Views of Light

The nature of light has intrigued philosophers and scientists for thousands of years. Because light travels in straight lines (*rectilinear propagation*), it was long believed that it was made of particles (corpuscles) that emanated from the source. However, some properties of light, such as the fuzziness produced when light passes around obstacles or through openings (*diffraction*), needed a different model of light where it behaved like a type of wave or propagating disturbance. In the early 1700s, both the particle and wave models of light prevailed with Sir Isaac Newton being a staunch supporter of the particle theory and others, such as the Dutch physicist, Christiaan Huygens, being proponents of the wave theory [2]. But, in the early 1800s, when the English physician Thomas Young showed that light diffracted and interfered to produce fringes when it passed through a double slit, the wave model of light became the accepted theory. At the end of the nineteenth century, the particle theory of light was revived by Einstein, who used particle behavior and a new type of theory – quantum physics – to explain the photoelectric effect. It is now widely accepted that light exhibits all three of these behaviors: particle, wave, and quantum, which are used to explain all known properties of light [2].

The history of light has also included interesting debates on how it is propagated. Ancient astronomers thought that light traveled at infinite speed. The Greek philosopher Aristotle proposed that light propagated through a hypothetical substance, luminiferous ether or light-bearing ether, because it was believed that all types of waves required a medium for propagation such as sound waves in air and water waves in water. However, light is very different from other types of waves. In 1881, in one of the most famous experiments with light, Michelson and Morley tried to measure the rotational motion of the earth in this

luminiferous ether [2]. However, they found that the light did not travel at different speeds for different paths. This failed experiment indirectly showed that light does not need a medium and can travel in free space (vacuum) and at constant speed. Michelson also carried out experiments to measure the speed of light by improving on Galileo's echo technique by replacing men carrying lanterns with a series of mirrors and a telescope. By adjusting the rotation speed of the mirrors, he could observe the light in the telescope, and from the rotation of the mirror and the known distance traveled by the light, he was able to determine the speed of light to about six significant figs [2]. Presently, the speed of light is the most accurately known fundamental constant of nature.

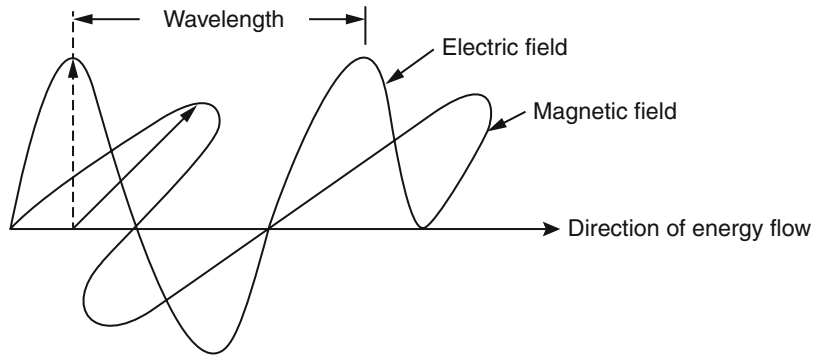
Description of Electromagnetic Radiation and Light

Electromagnetic radiation (emr) can also be described as the emission or transfer of energy from vibrating charged particles that creates a disturbance in the form of oscillating electric and magnetic fields (Fig. 1). The British physicist James Clerk Maxwell [3] was the first to discover that these electric and magnetic waves always occurred together, i.e., were coupled, giving rise to the name, *electromagnetic waves*. It was also found that the lines of force of these electric and magnetic waves were always perpendicular to each other and perpendicular to the direction of wave propagation, i.e., *transverse*, never longitudinal. This *transverse* nature of light gives rise to various phenomena associated with *polarization*.

Polarized electromagnetic radiation is defined as radiation whose electromagnetic field, which is transversal, is oriented in defined directions [1, 4] where this polarization can be linear, elliptic, or circular. Light is totally linearly polarized (or *plane polarized*) if the electric field vectors are all oriented in the same plane, parallel to a fixed direction which is referred to as the *polarization direction*. Light is *unpolarized* when its electric field vectors vibrate randomly in all directions; this can also be considered to be equal amounts of plane-polarized radiation whose

Light, Electromagnetic Spectrum,

Fig. 1 Simplified picture of an electromagnetic wave. The oscillations are perpendicular to each other and to the direction of energy flow



vibration directions are perpendicular to each another. The *state of polarization* of the radiation (linear, circular, and elliptic) is described by the phase relationship between these two orthogonal components. If this phase difference is zero or 180° , the radiation is *linearly polarized*; if the phase difference is 90° or 270° and both components have the same amplitude, the radiation is *circularly polarized*; and if the phase difference is not 0° , 180° , 90° , or 270° and/or the amplitudes are different, the radiation is *elliptically polarized*.

Light is *partially polarized* if the electric vectors have a preferred direction. If the electric vector is polarized vertical to the plane of incidence, this is referred to as *s-polarized* (or TE), and if the electric vector is polarized parallel to the plane of incidence, this is referred to as *p-polarized* or (TM).

In addition to its state of polarization, these *electromagnetic waves* are characterized by several other measurable properties, such as the number of oscillations per second, the separation between successive peaks or troughs, and the size or amount of oscillations. These quantities are described by the frequency, ν ; wavelength, λ ; and amplitude or intensity, I , respectively. The *intensity* describes the amount of energy flowing in the electromagnetic wave and is proportional to the square of the *amplitude*. It is also necessary to specify the *direction* that the wave is traveling, which is indicated in Fig. 1 by using an arrow. It can be seen that these waves crest up and down in a sinusoidal fashion, which is described as the *wave front* moving up and down. When more than one wave is traveling, it is important to describe how they differ in phase. Two waves

are *in-phase*, if their crests coincide and their troughs coincide. Two waves are *out-of-phase*, if the crest of one wave coincides with the trough of the second wave and vice versa. A *plane wave* is a wave whose surfaces of constant phase are infinite parallel lines transverse to the direction of motion. If all of the waves vibrate in-phase, the emr is referred to as *coherent*. The laser is an example of a coherent light source. If there is no fixed phase pattern between the waves emitted by a light source, the emr is referred to as *incoherent*. Most ordinary light sources, such as an incandescent lamp, are examples of incoherent light sources.

Properties of Electromagnetic Waves

As mentioned above, generally a wave, such as a water wave, is a propagating disturbance of some equilibrium state in a continuous medium. However, no medium is required for propagation of electromagnetic waves. This property of electromagnetic waves makes them unique from other types of waves that can be considered mechanical in nature and was first demonstrated by Einstein in 1905. He also proposed the idea that light could behave both as a wave and as a particle in order to explain the spectral distribution of the radiation emitted from a hot (incandescing) object. This type of electromagnetic radiation is referred to as *blackbody radiation*.

Electromagnetic waves of all wavelengths travel exactly at the same speed in vacuum. This fundamental constant of nature is known as the *speed of light in vacuum*, denoted by the symbol,

c. It has a value of 299,792,458-meter per second. The frequency, ν , is related to wavelength, λ , by $\nu = c\lambda^{-1}$. Waves are also sometimes described by their *wave number*, σ , which is the inverse of wavelength, defined by $\sigma = \lambda^{-1}$.

When an electromagnetic wave travels through other media such as air, glass, or water, its frequency remains constant, but its wavelength and speed are both reduced by a factor, n , the *refractive index* of the material. For commonly used optical glasses, the value of n ranges from 1.52 to 1.72 and is a function of frequency. For standard air, n has a value of 1.00028 [5] and can be ignored for most practical applications.

The behavior of electromagnetic waves at the boundary from one medium to another giving rise to reflection, transmission, and refraction phenomena is described by a set of wave equations known as *Maxwell's equations* [6].

These four equations have become the fundamental laws in electromagnetics connecting the principles of electricity and magnetism. Solving these equations reveals the wave equation and the form of the electromagnetic waves. The Maxwell's equations are not given here because this is outside the scope of this entry; the interested reader can find a good description in the following references [2, 6].

The smallest unit or quantity of electromagnetic radiation is called a *photon*. The energy of a photon is given by $E = h\nu$ where h is Planck's constant ($6.62607015 \times 10^{-34}$ Js). Light can then be described as a stream of individual photons, each with a definite energy and that can interfere with each other like waves and diffract around corners [3]. The motion of these photons is controlled by these same set of Maxwell wave equations [2, 3].

The *electromagnetic spectrum* spans the total range of wavelengths of electromagnetic radiation from the shortest to the longest wavelength that can be generated physically. This range of wavelengths spans practically from zero to near infinity and can be broadly divided into regions as shown in Table 1 [7], which includes radio waves, infrared, visible, ultraviolet, X-rays, and gamma rays. This division is not exact since there is a gradual transition from one region to the next, which is

Light, Electromagnetic Spectrum, Table 1 Regions of the electromagnetic spectrum

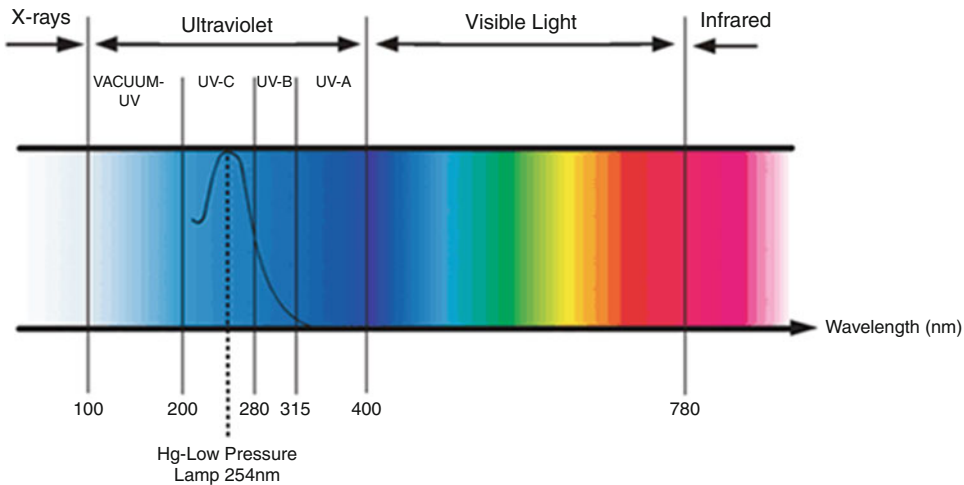
Wavelength range (nm)	Frequency range (s ⁻¹)	Description
<0.1 nm	10 ²⁰ –10 ²³	Gamma rays
0.1–10 nm	10 ¹⁷ –10 ²⁰	X-rays
10–400 nm	10 ¹⁵ –10 ¹⁷	Ultraviolet
400–700 nm	10 ¹⁴ –10 ¹⁵	Visible
700 nm to 1 mm	10 ¹¹ –10 ¹⁴	Infrared
1 mm to 1 cm	10 ¹⁰ –10 ¹¹	Microwaves
1 cm to 100 km	10 ³ –10 ¹⁰	Radio waves
100–1000 km	10 ² –10 ³	Audio frequency

shown schematically in Fig. 2 where the visible and ultraviolet regions are highlighted. This classification of light into regions is also shown as a function of frequency. It is important to note that the only difference between electromagnetic radiations in all these regions is its wavelength (and frequency). It has different descriptions because of the relationship between its frequencies and those that are excited in the various materials that the electromagnetic radiation can interact with. For example, the visual receptors in the human eye are only sensitive to electromagnetic radiation in a very narrow frequency range from 10¹⁴ to 10¹⁵ s⁻¹ (or 400–700 nm), whereas X-rays are used to excite features in the body that are of the size of an atom (0.1 nm). These different applications of emr are described in more detail below.

Interaction of Light with Matter

When light interacts with matter, different phenomena can occur depending upon the relationship of the wavelength (frequency) of the light with the physical size (resonant frequencies) of the interfering matter. This matter consists of atoms, ions, and molecules. As mentioned above, only light in the visible portion of the spectrum with frequencies from 10¹⁴ to 10¹⁵ s⁻¹ can stimulate the visual receptors in the human eye. Light can also exhibit wave or particle properties depending upon this relationship between

L



Light, Electromagnetic Spectrum, Fig. 2 The regions of the electromagnetic spectrum, highlighting the optical spectrum which includes the visible and ultraviolet regions

size (or frequency). For light waves that are closely spaced in relation to the spacing of the interfering matter, for example, a wave front passing through a slit or an opaque edge, secondary wave fronts are generated. These will interfere with the primary wave front, as well as with each other, and produce *diffraction* patterns. Since these secondary wave fronts were produced from the same primary wave, their phase will change in step. Thus, even for a normally incoherent light source, wave-like behavior such as *diffraction* and *interference* can be seen under certain conditions [8].

Many colorful effects are produced by this combination of light *diffraction* and *interference*. In nature, this can be seen, for example, in the beautiful array of colors displayed by mother of pearl (*opalescence*) or by thin film *iridescence*. For light waves that are very small in relation to the interfering particles, such as the interaction of visible light with dust particles or gas molecules in the atmosphere, *Rayleigh scattering* occurs which produces the blue color of the sky.

However, when light interacts with objects that are large in relationship to their wavelength, such as a brick wall or pane of glass, the light behaves more like a ray (particle) and the laws of geometrical optics can be applied.

Depending upon the transparency or translucency of the material and its surface quality, the light will be reflected, refracted, transmitted, absorbed, or scattered. The type of light interaction that dominates depends upon the precise nature of the matter, e.g., for a smooth opaque colorless solid, specular (mirror-like reflection) will dominate, whereas for an opaque colored material, diffuse reflection or scattering will dominate. The selective modification of the energy distribution of the incident light by the diffusely reflecting material gives rise to different colors.

Light Intensity Distribution and Color

The different wavelengths in the visible spectrum can also directly stimulate different colors in the human visual system. This dependence of color on wavelength is shown in Table 2.

Radiation of a single wavelength is called *monochromatic*. Except for lasers and certain specialized lamps, most sources of optical radiation emit energy over a broad wavelength region. The curve describing the power at each wavelength is called the *spectral power distribution* (SPD). A color stimulus generally has an SPD that varies with wavelength across the visible spectrum

producing a color or sensation that is a shade or mixture of the colors listed in Table 2. Stimuli containing all the visible wavelengths in roughly equal proportions appear white.

This is demonstrated by passing white light through a glass prism, which spreads the light out into a spectrum of colors (see Fig. 3).

However, white light can also be produced by superimposing discrete monochromatic lights. For example, white light can be produced by superimposing red, blue, and green lights in certain proportions. This was first demonstrated by Thomas Young [3] who used this experimental finding to theorize that the human eye contained three different types of color receptors, which is known as the trichromatic theory of human color vision.

Because of this trichromacy, white light can be produced by many different SPDs. The variation in these SPDs is illustrated in Fig. 4, which shows the SPD of a typical tungsten lamp, a fluorescent lamp, a white LED source produced by a blue LED and yellow phosphor, and noon daylight. The most prevalent source of light in our natural environment – the sun – has an SPD that can

change dramatically during the course of the day, from a bluish cast at noon to a reddish-orange cast at sunset. These variations in SPD of the light source are very important for visual evaluations of colored goods. For this reason, the SPDs of different phases of daylight have been standardized by the International Commission on Illumination (CIE) as a series of D illuminants. Shown in Fig. 4 is the SPD of CIE standard illuminant D65, which is average daylight with an approximate correlated color temperature of 6500 K.

Applications of Electromagnetic Radiation

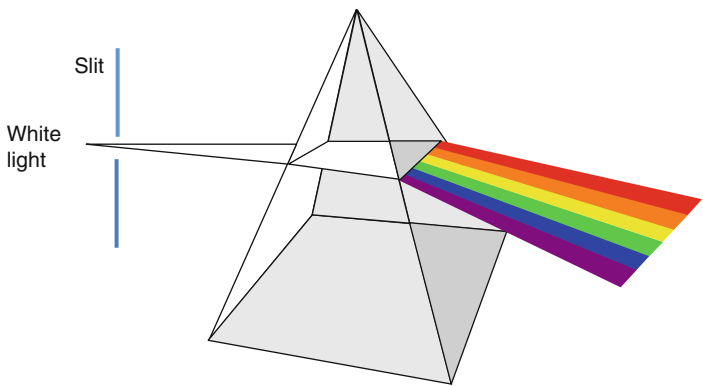
The applications of electromagnetic radiation (emr) cover almost all aspects of our daily lives. The importance of visible light for human vision has already been discussed. Light emanating from the sun spans the solar wavelength region from about 250 nm in the UV to 2500 nm in the near-infrared region. This solar radiation is extremely important to the energy balance of the earth and for various building design applications, such as the design of glazing materials. It is also exploited in light-based technologies, such as photovoltaics.

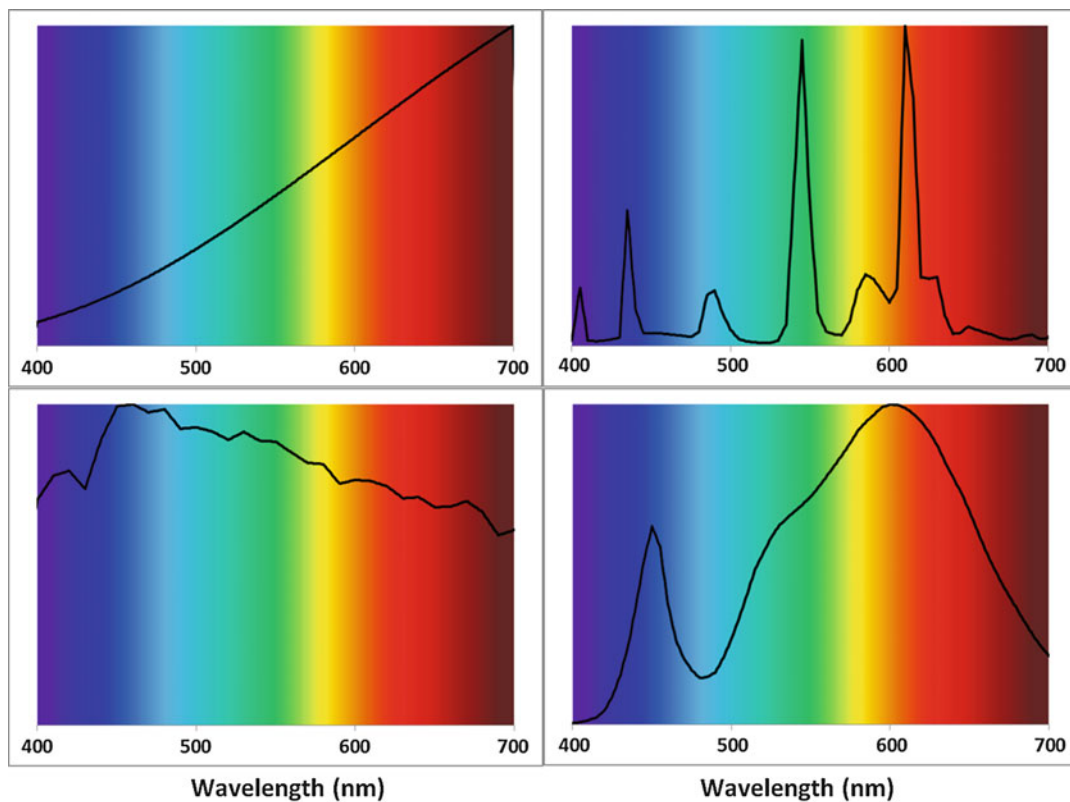
Light in the infrared region of the spectrum is felt as heat and is used in the design of radiant heating, solar panels, and collectors. The frequencies in the infrared region are close to the resonant vibrational frequencies in molecules and can be used for molecular identification by spectroscopic means. Similarly, the frequencies in the

Light, Electromagnetic Spectrum, Table 2 Dependence of color on wavelength

Wavelength range (nm)	Color
400–430 nm	Violet
430–480 nm	Blue
480–560 nm	Green
560–590 nm	Yellow
590–620 nm	Orange
620–700 nm	Red

Light, Electromagnetic Spectrum, Fig. 3 Use of a prism to separate white lights into a spectrum of colors





Light, Electromagnetic Spectrum, Fig. 4 Relative spectral power distributions (SPDs) of four different sources of white light; CIE illuminant (if applicable) given in

parentheses: top left: tungsten lamp (CIE A); bottom left: noon daylight (CIE D65); top right: fluorescent lamp (FL12); and bottom right: blue LED + yellow phosphor

microwave region are close to the resonant rotational frequencies in molecules, in particular the water molecule. Since most foods contain a significant amount of water, the microwave radiation heats the food by the increased rotational motion of the interspersed water molecules. These microwaves pass through glass, ceramic, and plastic but reflect from metal which can also be used to advantage in the selection of the food container and walls of the microwave, respectively.

Light in the ultraviolet (UV) region has many important uses. In the early nineteenth century, it was called “chemical rays” because it was found that UV radiation could cause chemical changes. This was used to advantage to harden special glues in inks, coatings, and adhesives more quickly, in a process called UV curing through polymerization. The skin of our bodies needs exposure to UV with wavelengths in the region

280–315 nm (UVB) in order to stimulate the production of vitamin D. However, too much exposure to UVB can cause skin cancers, so it is important to find the optimum exposure level. UV light with wavelengths in the region 200–280 nm (UVC) is called “germicidal UV” because it has sufficient energy to inactivate bacteria and viruses by disabling their DNA strands. For this reason, it is also used for sterilization of surfaces such as medical equipment. This type of UVC radiation is often produced by a low-pressure mercury vapor lamp (see Fig. 2). UV radiation is also needed to excite a class of molecules referred to as fluorescent-whitening agents (FWAs) or optical brighteners in manufactured white goods, such as paper, fabrics, detergents, soaps, and cosmetics, which emit in the blue portion of the visible spectrum and produce an enhanced whiteness effect. UV radiation is also of the appropriate frequencies

to absorb ozone molecules in the atmosphere so very little UV radiation reaches the earth's surface.

X-rays have a higher frequency than UV radiation and can pass through the skin and soft tissue, but they do not easily pass through the bone or metal. This type of emr is used to produce photographs of bones in medical diagnostics to check for damage such as fractures. The frequencies of gamma radiation, on the other hand, are more penetrating and can destroy chemical bonds by interacting with the electrons of the constituent atoms or disrupt DNA bonds, resulting in prevention of cellular division. This can be used to advantage to kill microorganisms throughout the material. Gamma sterilization is used, for example, in sterilization of water and food products.

Other important applications of emr are radio, television, and electric current.

Cross-References

- [Blackbody and Blackbody Radiation](#)
- [Iridescence \(Goniochromism\)](#)
- [Rayleigh and Mie Scattering](#)
- [Spectral Power Distribution](#)

References

1. CIE S 017/E: ILV – International Lighting Vocabulary, 2nd edn. CIE Central Bureau, Vienna (2020)
2. Jenkins, F.A., White, H.E.: *Fundamentals of Optics*, 4th edn. McGraw-Hill, New York (1976)
3. Overheim, R.D., Wagner, D.L.: *Light and Color*. Wiley, New York (1982)
4. Shurcliff, W.E.: *Polarized Light: Production and Use*. Harvard University Press, Cambridge, MA (1962)
5. Ciddor, P.E.: Refractive index of air: new equations for the visible and near infrared. *Appl. Opt.* **35**, 1566–1573 (1996)
6. Stern, F.: In: Seitz F., Turnbull D. (eds.): *Elementary Theory of the Optical Properties of Solids*. Solid State Physics, vol. 15, pp 301–304. Academic, New York (1963)
7. Williamson, S.J., Cummins, H.: *Light and Color in Nature and Art*. Wiley, New York (1983)
8. Falk, D., Brill, D., Stork, D.: *Seeing the Light: Optics in Nature, Photography, Color, Vision, and Holography*. Harper & Rowe Publishers, New York (1986)

Light-Emitting Diode, LED

Wout van Bommel
van Bommel Lighting Consultant,
Nuenen, The Netherlands

Synonyms

[Optoelectronic light source](#); [Solid-state light source](#)

Definition

A light source that produces light as a result of recombination of positively charged holes and negatively charged electrons at the junction of inorganic, solid-state, p- and n-type semiconductor material.

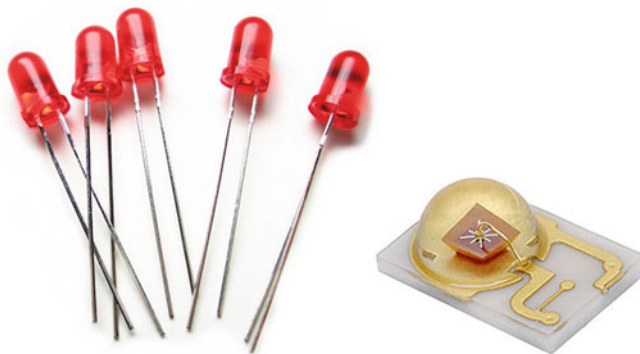
Solid-State LED Light Sources

LEDs are solid-state radiators where the light is created inside solid-state material [1–3]. Common semiconductor diode chips, used today in so many electrical circuits, all use much the same technology. The light-radiating diode versions are called light-emitting diodes or LEDs. The name LED is commonly used for light-emitting diodes made of inorganic semiconductor material; they are point light sources. Light-emitting diodes made of organic semiconductor material are referred to as OLEDs; they are planar light sources. A separate chapter deals with OLEDs.

Until the mid-1990s of the last century, LEDs had a low lumen output and low efficiency, making them only suitable as small indicator lamps (e.g., in electrical appliances). Today, the efficacy of LEDs is comparable to that of gas discharge lamps. The lumen output of a single LED can be more than 1000 lm. To distinguish these LEDs from the indicator type of LEDs, they are referred to as high-brightness or high-power LEDs (Fig. 1). Further improvements in high-brightness LEDs are expected, ultimately leading to

Light-Emitting Diode, LED,

Fig. 1 Indicator lamp LEDs (*left*) and a high-power LED (*right*)



efficacies of probably slightly more than 200 lm/W (for white-light LEDs). This is approximately twice the efficacy of today's most efficient white-light gas discharge lamps. With its light-emitting surface of some 0.5–5 mm², an individual LED chip represents the smallest artificial light source. LEDs have a long lifetime and are, given the solid-state material, extremely sturdy. They are available in white and in colored-light versions. The colored versions are extensively used in traffic signs. Colored versions were also the first ones to be used on a large scale for lighting: specifically, the exterior floodlighting of buildings and monuments. Both the efficacy and the color quality of white LEDs have been improved so much that they are now used in many different lighting applications, including road lighting, indoor accent lighting, domestic lighting, and automobile lighting. Examples of the use of LEDs for office lighting and outdoor sports lighting can also be found. Given the potential for a further increase in their efficacy, the number of LED applications is bound to increase. Their small size and their availability in many different colors and the easiness of lighting control, both in terms of dimming and color changing, are properties that permit of entirely new applications.

Working Principle

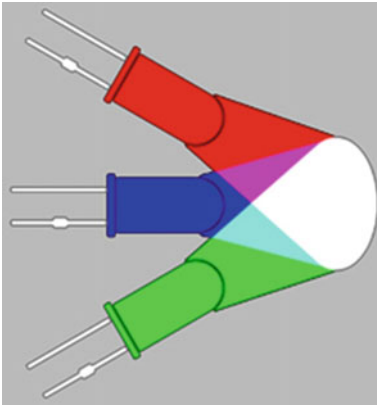
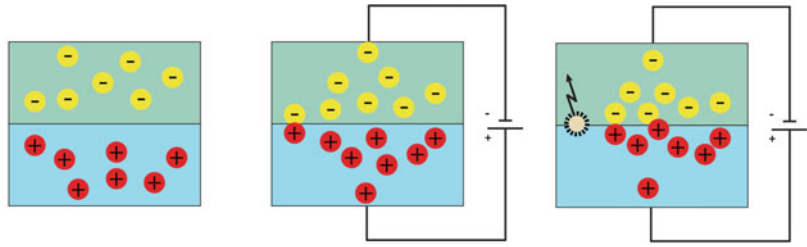
Principle of Solid-State Radiation

Like any diode, an LED consists of layers of p-type and n-type of semiconductor material. The n-type of material has an excess of negatively

charged electrons, whereas the p-type material has a deficiency of electrons, viz., positively charged holes. Applying a voltage across the p-n semiconductor layer pushes the n- and p-type atoms toward the junction of the two materials (Fig. 2). Here the n-type of atoms “donate” their excess electrons to a p-type of atom that is deficient in electrons. This process is called recombination. In doing so, the electrons move from a high level of energy to a lower one, the energy difference being emitted as light. The wavelength of the light is dependent on the energy-level difference between the p and n materials, which in turn depends on the semiconductor material used: different semiconductor materials emit different wavelengths, and thus different colors, of light. The process is very much like the process of excited electrons in a gas discharge falling back to their original orbit with a lower energy level while emitting light. In this sense, it is surprising that the expression “solid-state discharge” and “solid-state discharge lamp” is hardly ever used. Not all recombinations result in light emission. Some recombinations are non-radiative and just heat up the solid material. This limits the efficiency of light creation. A further limitation of the efficiency is caused by absorption of light in the solid material of the chip itself. Improvements in radiative-recombination efficiency and light-extraction efficiency have been the most important reasons for the dramatic improvements in efficacy of LEDs during the past decade. Further improvements will be sure to further significantly increase the efficacy of LEDs over the coming decade.

Light-Emitting Diode, LED,

Fig. 2 Principle of operation of solid-state radiators



Light-Emitting Diode, LED, Fig. 3 White light by combining red, green, and blue LED light [4]

Like gas discharge lamps, solid-state lamps cannot function when they are operated direct from the mains supply voltage. Solid-state light sources are low-voltage rectifiers that allow current to pass in one direction only. This means that the AC mains supply has to be transformed to low voltage and then rectified into a DC supply. Small fluctuations in supply voltage cause large variations in current that can damage the light source. High-power LEDs need an electronic driver to obtain a constant current characteristic.

Principle of White LEDs

The spectrum of a single LED is always narrow. Consequently, its light is colored. White LED light can nevertheless be obtained by combining three (or more) different colored LED chips. A common method is to combine red, green, and blue LED chips into a single module to produce white light (RGB LED, Fig. 3). Sometimes the three chips are mounted in the same LED package (Fig. 4). The color rendering of an



Light-Emitting Diode, LED, Fig. 4 RGB chips mounted in the same LED package

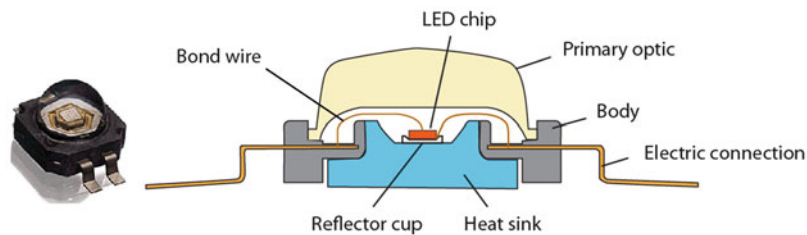
“RGB white-light” system is not good, since large areas of the full-color spectrum are not included in its light. Sometimes an amber color chip is added to the RGB combination to improve the color quality of the light (RGBA LED). Research is going on to produce single, multilayer LED chips, each layer producing a specific color of light. A single LED producing red, green, and blue light would, therefore, result in white light.

Good-quality white light, which is especially important when it comes to providing good color rendering, is obtained by using a blue LED chip in combination with a fluorescent material that converts much of the blue light into light of different wavelengths spread over almost the whole visible spectrum. In LED technology it is customary to call such fluorescent materials, phosphors: hence white LEDs based on this principle are called “white-phosphor LEDs.” By mixing different phosphors in different proportions, white LEDs producing different tints of white light with different color-rendering capabilities can be made.

Materials and Construction

The LED chip is embedded in a larger structure for mechanical protection, for the electrical

Light-Emitting Diode, LED, Fig. 5 The main components of a high-brightness LED [4]



InGaN Colours					AlInGaP Colours			
								
450 nm Blue	498-500 nm Green-Blue	505 nm Blue-Green	525 nm Green		590 nm Amber	605 nm Orange	615 nm Red-Orange	626 nm Red

Light-Emitting Diode, LED, Fig. 6 The main semiconductor material elements used in high-brightness LEDs, with examples of the corresponding light colors [4]

connections, for thermal management, and for efficient light out-coupling. The main parts of an LED are (Fig. 5):

- Semiconductor LED chip
- Reflector cup
- Supporting body
- Electrodes and bond wires
- Heat sink
- Primary optics
- Phosphors (for white LEDs)

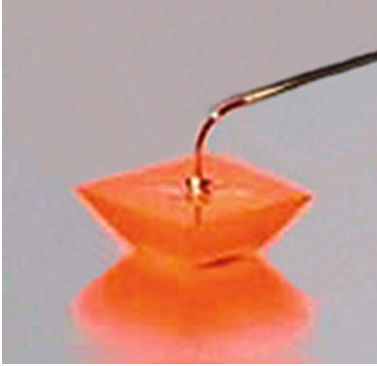
Semiconductor Chip Material

The p-n semiconductor sandwich forms the heart of the LED and is called the LED chip or die. The semiconductor material used determines the wavelength and, thus, the color of the light emitted. For LEDs, compound semiconductor material is used, which is composed of different crystalline solids. These are doped with very small quantities of other elements (impurities) to give their typical n and p properties. For the colors blue, green, and cyan, the elements indium, gallium, and nitride (InGaN) are used in different compositions (Fig. 6, left). The elements aluminum, indium,

gallium, and phosphide (AlInGaP) are used to produce the colors amber, orange, and red (Fig. 6, right). It was the Japanese Nakamura who, in 1993, succeeded for the first time in producing a blue LED suitable for mass production. This was the “missing link” that enabled the production of white-phosphor LEDs and the production of white LED light on the basis of mixing the light of red, green, and blue LEDs. As can be seen from Fig. 6, today a very small area of the spectrum, in greenish-yellow, is still missing.

Shape of the Chip

Unfortunately, the LED chip is a “photon or light trap”: that is to say, much of the light emitted within the chip is internally reflected by its surfaces (borders between the material and air) and, ultimately, after multi-reflections, absorbed in the material (heating up the material). Only light that hits the outer surface more or less perpendicularly (approximately 20°) can leave the material. By giving the chip a specific shape and by keeping it thin, the so-called light-extraction efficiency can be improved. Figure 7 gives an example of such a specially shaped chip.



Light-Emitting Diode, LED, Fig. 7 An example of a specially shaped LED chip that improves light-extraction efficiency. On top of the chip, the anode with its bond wire can be seen [4]

Reflector Cup

In many cases, the chip is placed in a reflector cup which, because of its shape, helps to direct the light in an upward direction. Highly reflective material is used, for example, metal or ceramic material.

Primary Optics

The silicon lens on top of the LED chip serves as protection for the chip. More importantly, it helps in increasing the light extraction from the chip and as such is essential for a high lumen efficacy of the LED. This is because by introducing a lens medium between the chip and the air with a refractive index value between that of air and that of the chip material, the angle over which light can escape from the chip increases.

Electrodes and Bond Wires

In order to be able to apply power to the chip, the p and n parts of the chip have metal contacts called electrodes. Bond wires connect the electrodes with the electrical connections. They are usually gold wires. Since the electrodes intercept light leaving the chip, the dimensioning of the electrodes and bond wires, especially on the side of the main light-escape route, is one of the factors that determines the light efficiency of the LED.

Heat Sink

LEDs do not radiate infrared radiation and consequently give a cool beam of light. However, this

does not mean that they do not generate heat. Non-radiative recombinations of electrons and holes in the p-n sandwich and light trapped in the chip do heat up the chip. The larger the power of the chip and the lower its luminous efficacy, the higher is this heating effect. The higher the temperature of the p-n junction in the chip, the lower the light output of the chip. A too-high chip temperature also severely shortens LED life, and it also slightly shifts the emitted wavelength and thus the color of the LED. Effective thermal management is therefore critical for the proper functioning of LEDs. All high-power high-brightness LEDs, therefore, have a heat sink of high-thermal-conductivity material (like aluminum or copper) on their rear side to conduct the heat away from the chip toward the outside world through the luminaire housing. LED luminaires must, therefore, incorporate in their design thermal conduction and convection features (such as cooling fins, Fig. 8) to dissipate the heat to the immediate surroundings. For retrofit LED lamps (LED bulbs), the size of the heat sink is limited by the size of the bulb. The heat sink, therefore, has a limited capacity, thus limiting the power of the retrofit LED bulbs.

Phosphors

As mentioned above, the most important method for producing white light with LEDs is by applying a phosphor coating to a blue-light LED that converts part of the blue light into longer-wavelength, green, yellow, and red light. Different compositions of different phosphors are used to produce white light of different color tints. Since the basis of the process is the blue light of the chip, the final efficacy will become higher as more blue is kept in the light. However, this implies a high color temperature (cool-white light) and relatively poor color rendering. If with a different phosphor composition, more blue light is converted, the color quality will improve at the expense of a somewhat lower efficacy.

Phosphors Applied Direct to the Chip The phosphor is often applied on or very near to the blue LED chip (Fig. 9). The thickness of the layer has to be very uniform as a variation in thickness

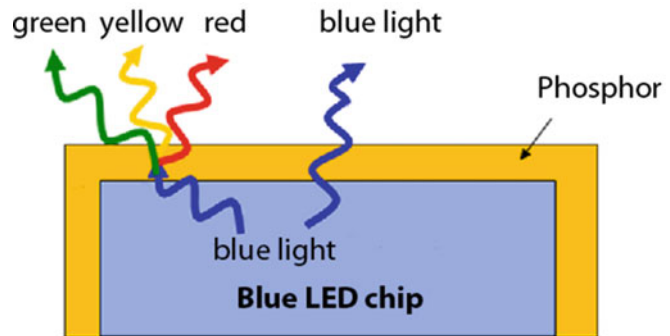
Light-Emitting Diode, LED,

Fig. 8 Examples of luminaires with cooling fins



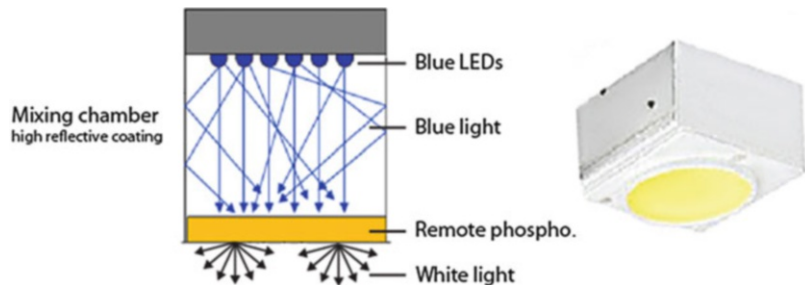
Light-Emitting Diode, LED,

Fig. 9 Principle of creating white light with a blue-light chip covered with phosphor



Light-Emitting Diode, LED,

Fig. 10 Principle of a remote-phosphor LED module creating white light [4]



of the phosphor layer causes a variation of the color temperature in the light beam.

Remote Phosphor In the case of multi-LED units, the phosphor is sometimes applied at a greater distance from the LEDs. Such modules are called remote-phosphor LED modules (Fig. 10). Here, a number of blue LEDs are placed inside a mixing chamber of high and diffuse reflective material. The phosphor layer, positioned remotely from the LEDs on the bottom of the chamber, converts the blue light of the chips into white light. In this way, thanks to the mixing process, small differences in light output and/or color of individual chips are not visible. The risk

for disturbing glare is also reduced because the light intensity from the large-sized phosphor layer is much lower than the intensities of small, individual LEDs. In regular phosphor LEDs, the hot LED heats up the phosphor that is applied on the LED itself, reducing its conversion efficiency. Heating of remote positioned phosphor is much less, and thus the conversion efficiency of remote-phosphor systems is higher. In regular phosphor LEDs, a relatively large amount of light from the LED is reflected back from the phosphor toward the LED and absorbed there. In the remote-phosphor situation, most of the light reflected back from the phosphor layer takes part in the mixing process without reaching the LEDs, thus



Light-Emitting Diode, LED, Fig. 11 LED cluster module with multiple LEDs mounted on a printed circuit board

further increasing the final efficiency of the system. The phosphors used for the blue-light conversion appear yellow when they are not activated, that is to say, when the LED is not switched on (see Fig. 10 right). This sometimes makes people think, erroneously, that the blue light is filtered through a yellow filter. It is really wavelength conversion and not light filtering that takes place.

LED Cluster Modules

The luminous flux of one individual LED is quite low compared to that of most conventional light sources. Multiple LEDs are therefore often mounted on a printed circuit board (PCB) to obtain an LED module emitting a high luminous flux (Fig. 11). The PCB establishes the electrical connections between all components and the external electrical driver. The PCB must also conduct the heat from the heat sinks of the LEDs to the outside world.

Properties

Temperature of the Chip Junction

It has already been mentioned that with rising temperature of the p-n chip junction, the

performance of LEDs decreases: particularly the light output and lifetime. The performance data are usually specified for a junction temperature (T_j) of 25 °C. However, under normal operating conditions, a junction temperature of 60–90 °C is easily obtained. Depending on the LED type, the lumen output falls to 60–90% when the junction temperature increases from 25 °C to 80 °C. Amber and red LEDs are the most sensitive to changes in junction temperature and blue LEDs the least.

Binning

The mass production of LEDs results in LEDs of the same type varying in color, light output, and voltage. In order to ensure that LEDs nevertheless conform to specification, LED manufacturers use a process called binning in their production process. At the end of the manufacturing process, LED properties are measured, and LEDs are subsequently sorted into subclasses or “bins” of defined properties. As far as the color quality is concerned, the tolerances in these definitions (based on MacAdam ellipses) are such that visible differences in color between LEDs from the same bin are minimized. As the definition of a bin does not change with time, the same quality is also assured from production run to production run.

With the advancement of knowledge concerning LED materials and the mass production process, it may be expected that binning will ultimately no longer be required (“binning-free LEDs”).

Energy Balance

The energy balance of an LED is much easier to specify than that of conventional light sources. This is because no energy is radiated in the UV and infrared region of the spectrum, which means that the energy balance comprises only visible radiant energy and heat energy. Today, white LEDs transform 20–30% of the input power into visible light and the remaining part into heat. The light percentage comes close to that of fluorescent lamps and will soon supersede it.

System Luminous Efficacy

Sometimes, luminous efficacies are specified for the bare chip. It is evident that the ancillary

devices described in the previous sections, which are essential for the proper functioning of LEDs, absorb light. The only sensible thing to do, therefore, is to specify luminous efficacy (and light output as well) for the total LED package. As with most conventional lamps, the luminous efficacy of LEDs is dependent on the power of the LED and on the color quality of the light it produces. Higher-power LEDs have higher efficacies, while those with better color rendering have lower efficacies. Today cool-white LEDs are commercially available in efficacies up to some 150 lm/W and warm-white LEDs with color-rendering indices larger than 80 in efficacies of around 130 lm/W (all lm/W values include driver losses). Retrofit LED bulbs with warm-white light (around 2700 K) and color-rendering index better than 80 are available in efficacies up to 80 lm/W. Here, too, cooler-white versions are slightly more efficient.

As discussed at the beginning of this chapter, in coming years, it may be expected to see further important improvements in luminous efficacies for white LEDs – even up to slightly more than 200 lm/W.

Lumen-Package Range

Today, single LEDs exist in lumen packages varying from a few lumens (indicator lamps) to more than 1000 lm. In the latter case, severe screening is called for to restrict glare, because so much light comes from such a very small light-emitting

surface. By mounting multi-LEDs on a printed circuit board, LED modules can be achieved with much larger lumen packages.

Color Characteristics

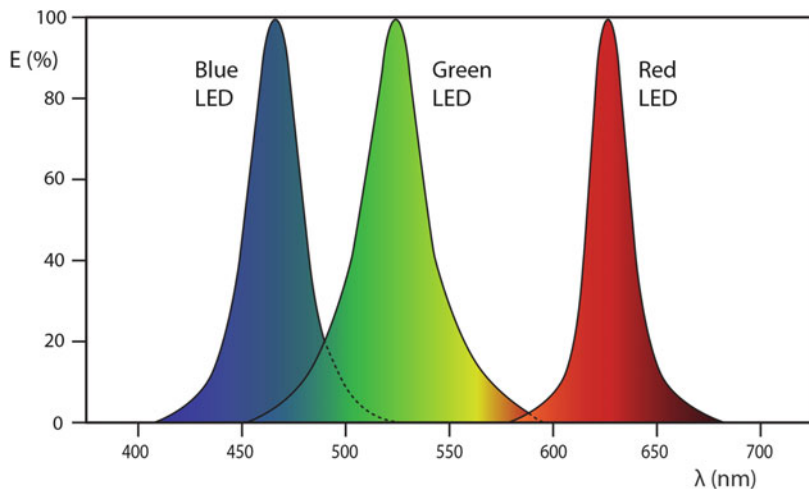
Principally, LEDs have a quasi-monochromatic, narrow-band spectrum. Figure 12 shows the spectra of blue, green, and red LEDs. The half-maximum width of the spectrum bands is smaller than ca. 50 nm. It has already been shown what colors can be produced by using different semiconductor materials (see Fig. 6). With some LED modules, different colored LEDs can be controlled (dimmed) individually. With such LED modules (e.g., making use of RGB LEDs), the color of the light can be dynamically changed from white to all colors of the spectrum.

White-phosphor LEDs can have a near-continuous spectrum (Figs. 13 and 14). By applying different phosphors, white light within the color temperature range of 2700–10,000 K can be produced. Often, the higher-color-temperature versions have only moderate color rendering (R_a between 50 and 75). In the lower-color-temperature versions, LEDs are available with good (R_a larger than 80) to excellent color rendering (R_a larger than 90 or even 95).

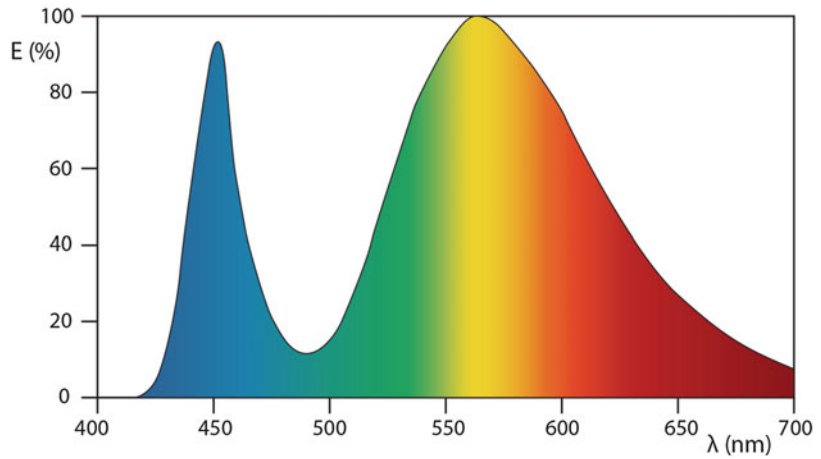
Beam Control

For lighting designers, one of the most exciting properties of an LED is its small light-emitting surface. This allows the creation of very

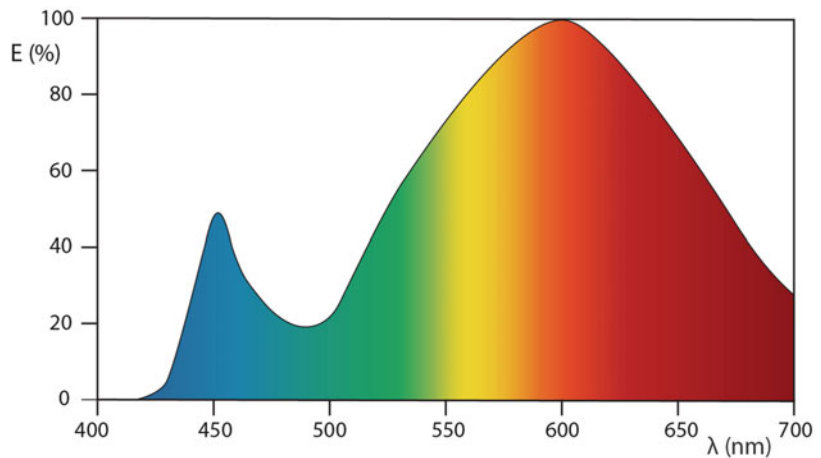
Light-Emitting Diode, LED, Fig. 12 The relative spectral power distributions of a typical blue, green, and red LED



Light-Emitting Diode, LED, Fig. 13 The relative spectral power distribution of a white-phosphor LED with color temperature $T_k = 4000$ K and color rendering $R_a = 70$



Light-Emitting Diode, LED, Fig. 14 The relative spectral power distribution of a white-phosphor LED with color temperature $T_k = 2750$ K and color rendering $R_a = 85$



L

accurately defined beams. As an illustration of this, Fig. 15 shows a near-parallel light beam made with an LED-line luminaire that is impossible to create with conventional light sources. The other side of the coin is that small light-emitting surfaces often need professional screening in order to limit excessive glare. Multi-LED luminaires also have multi-light beams that may cause multiple shadows because a lighted object is illuminated from many slightly different directions. With RGB color mixing, this may lead to disturbing, multicolored shadows.

Lifetime

In the case of high-performance LEDs, it takes a very long time before they actually fail – usually considerably more than 50,000 h. Before that

time, however, their lumen depreciation is so significant that the LED is no longer giving sufficient light for most applications. Therefore, for LED lifetime specifications, the length of time that it takes to reach a certain percentage of its initial light value is used. Based on a depreciation value of 70%, lifetime values of between 35,000 and 50,000 h are common for high-performance LEDs. LED bulbs, with their limited space for handling heat, have a lifetime of some 25,000–35,000 h (25–35 times longer than an incandescent lamp).

Lumen Depreciation

The electric current passing through the chip's junction, and the heat generated in it, degrades the chip material and is so responsible for light

depreciation. Decoloration of the housing and yellowing of the primary lens may be further reasons for the lumen depreciation. In white-phosphor LEDs, chemical degradation of the

phosphor material also causes lumen depreciation. As mentioned above in the section on LED lifetime, for high-performance LEDs, 30% lumen depreciation is reached at around 35,000–50,000 h.



Light-Emitting Diode, LED, Fig. 15 A near-parallel light beam with an LED-line luminaire [4]

Run-Up and Reignition

LEDs give their full light output immediately after switch on and after reignition.

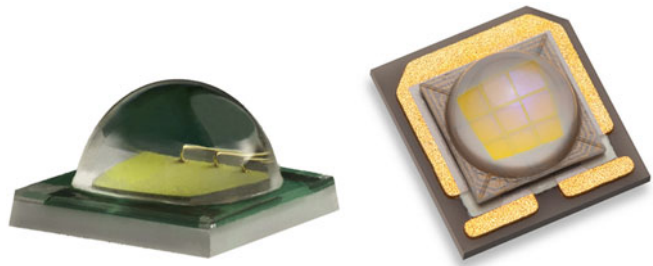
Dimming

LEDs can be dimmed by simple pulse-width modulation down to 5% of full light. Not all retrofit LED lamps can be dimmed on regular, commercially available dimmers. Special retrofit LED lamps that are designed to be dimmed on such dimmers are so specified on their packaging.

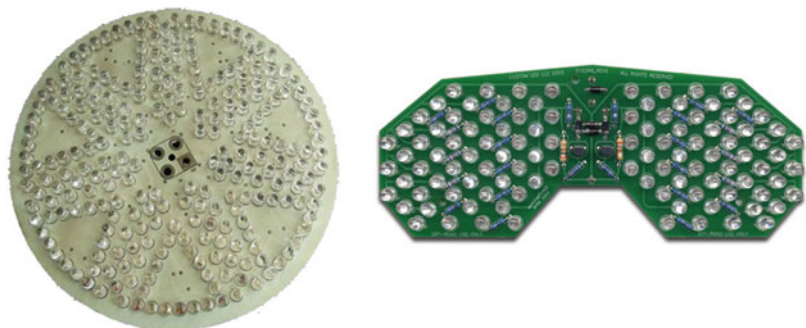
Ambient-Temperature Sensitivity

As has already been mentioned several times, limitation of the junction temperature of the LED chip is essential for the proper functioning of LEDs in terms of lumen output, lumen efficacy, lamp life, and even color properties. In high-temperature environments, the products perform worse, while at low temperature they perform

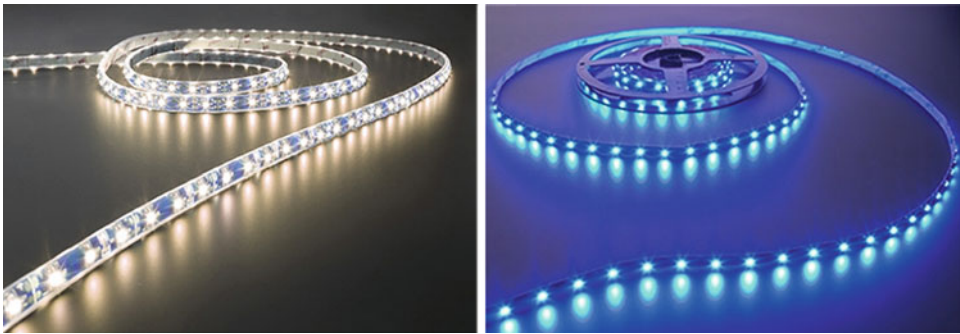
Light-Emitting Diode, LED, Fig. 16 Single LEDs



Light-Emitting Diode, LED, Fig. 17 PCB-mounted LEDs



Light-Emitting Diode, LED, Fig. 18 LED engines (*bottom*: for use in road lighting)



Light-Emitting Diode, LED, Fig. 19 Foldable LED string

L

Light-Emitting Diode, LED, Fig. 20 Retrofit LED lamps for incandescent, halogen, and fluorescent lamps, respectively



Light-Emitting Diode, LED,

Fig. 21 Designer type of retrofit LED lamps



better. The actual influence of the junction temperature is different for the different types of LEDs. In extreme temperature environments, therefore, relevant information for a particular product has to be obtained from the manufacturer.

Mains Voltage Variations

LED drivers are designed to drive LEDs on constant current. In this way the influence of mains voltage variations is not an issue.

UV and IR Component

LEDs radiate only visible radiation. There is no ultraviolet or infrared radiation.

Product Range

LED products are available as:

- Single LEDs (Fig. 16).
- Multiple LEDs on flat or three-dimensional PCB boards (Fig. 17).
- LED modules (or LED engines) with secondary optics and with or without built-in driver that can be used in the same way as lamps. Interfaces are standardized for interchangeability (Fig. 18).
- Multiple LEDs on strings (Fig. 19).
- Retrofit LED lamps with built-in driver and conventional design (Fig. 20); new types of designs are introduced for retrofit LED lamps as well (Fig. 21).

Cross-References

- [Light-Emitting Diode, OLED](#)
- [Phosphors and Fluorescent Powders](#)

References

1. Schubert, E.F.: *Light Emitting Diodes*, 2nd edn. Cambridge University Press, Cambridge, UK (2006)
2. Mottier, P.: *Led for Lighting Applications*. Wiley, Hoboken (2010)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *IES Handbook*, 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)
4. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course Book. Philips Lighting, Eindhoven (2012)

Light-Emitting Diode, OLED

Wout van Bommel
van Bommel Lighting Consultant, Nuenen, The Netherlands

Synonyms

[Optoelectronic light source](#); [Solid-state light source](#)

Definition

A light source that produces light as a result of recombination of positively charged holes and negatively charged electrons at the junction of organic, solid-state, p- and n-type planar semiconductor materials.

Solid-State Light Sources

LEDs are solid-state radiators where the light is created inside solid-state material [2–4]. The light-radiating diode versions are called light-emitting diodes or LEDs. Light-emitting diodes made of organic semiconductor material are referred to as OLEDs; they are planar light sources (Fig. 1). Organic semiconductor material is a chemical compound whose molecules contain carbon and hydrogen (C-H bonds). A property of carbon is that it can link with other carbon atoms to form long chains, a process called polymerization. The chains are called polymers. Artificially (synthetically) produced polymers are called plastics. Semiconductor polymers can be produced as thin films with a thickness of 10 to 100 nm. They combine the electric properties of semiconductors with the mechanical and chemical properties of plastics. These semiconductor films can be used to make relatively large-sized planar light sources of relatively low luminance with an inherently diffuse light distribution.

Serious development of commercial OLEDs only started around the year 2000. Initially,

expectations were that a relatively fast increase in luminous efficacy to more than 100 lm/W and in dimensions of much more than 1 x 1 m would be possible. Expectations today are that developments will be distinctly slower (Fig. 2). OLEDs can be produced to give most colors of the spectrum and white. The technology permits producing OLED windows that are transparent when not switched on (Fig. 3). Currently, an important application for OLEDs is displays for mobile phones and other display screens including television screens. Today, the most interesting lighting application for OLEDs is decorative lighting. For general lighting purposes, the efficacies are still too low.

Working Principle

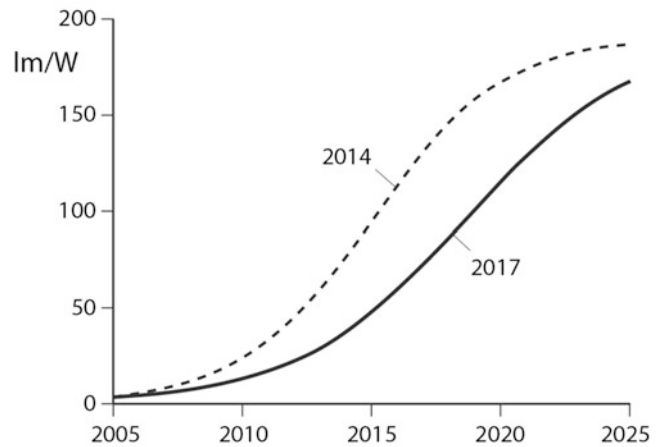
The process that is responsible for the emission of light in OLEDs is very similar to that of inorganic LEDs: positively charged holes and negatively charged electrons are pushed through semiconductor layers toward each other and recombine (see the entry ► “[Light-Emitting Diode, LED](#)”). Part of these recombinations results in the emission of light. The color of the light is dependent on the composition of the semiconductor material. White light can be obtained by combining layers that emit different wavelengths (including red, green, and blue) or by bringing phosphorescent materials or quantum dots in the emissive layers. The number of layers varies and is dependent on the desired spectrum and the technology used.

Light-Emitting Diode, OLED, Fig. 1 Flat OLED light sources



Light-Emitting Diode, OLED,

Fig. 2 OLED panel efficacy projections made in 2014 and 2017 by the US Department of the Environment, DOE [5], respectively, for commercial products, based on best-in-class performance



Light-Emitting Diode, OLED,

Fig. 3 OLED window [1]



Materials and Construction

The semiconductor organic layers are placed between electrodes, the one on the light-escape side being transparent (Fig. 4). The layers are supported by a glass substrate and, for protection of the organic materials against oxygen and water, are sealed in glass. Development is going on to substitute the glass seal by thin-film encapsulation. This development not only reduces the thickness and weight of OLEDs but also opens up the possibility of bendable OLEDs. The first laboratory examples have already been produced.

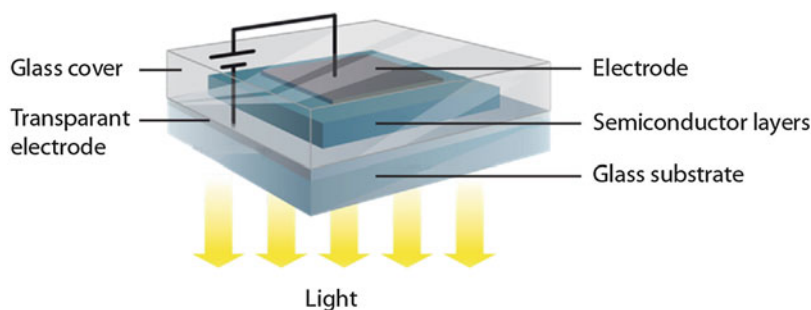
Properties

System Luminous Efficacy and Brightness

Commercially available white OLEDs now have efficacies around 55 lm/W. Ultimately, white-light, large-sized OLEDs with efficacies up to 150 lm/W would seem to be possible (see Fig. 2). Products with luminance values of up to 3000 cd/m² have been shown. Compare this with a luminance value of some 6000–10,000 cd/m² of fluorescent tubes.

Light-Emitting Diode, OLED,

Fig. 4 Composition of an OLED [1]



Dimensions

The OLED panel can be as thin as 1 mm. Planar dimensions of commercially available OLED panels for lighting purposes go up to some 1×1 m.

Lifetime

Based on a lumen maintenance value of 70% (L70), lifetime values of high-quality OLEDs have reached 30,000 hrs, while values of 50,000 hrs have been reported.

Color Properties

The quality of white-light OLEDs depends on the combination of the semiconductor and phosphorescent materials used. Spectra similar to those of inorganic LEDs are obtained (see Fig. 11.32). As with LEDs the CCT range of OLEDs goes from 2500 K to 10,000 K. Color rendering varies from poor to excellent.

Light Distribution

The light distribution of an OLED panel is inherently diffuse (Lambertian distribution) and its luminance uniform. Light beam shaping techniques, possibly integrated with light extraction improving technologies, making use of, for example, microlenses or layers of optical films applied on top of the panel, have been described.

Run-Up and Reignition

OLEDs, like LEDs, give their full light output immediately after switch-on and after reignition.

Dimming

OLEDs can be dimmed by pulse-width modulation (PWM dimming) or by reducing the drive

current (analogue dimming). The advantages and disadvantages of both methods are described in the next section.

Ambient-Temperature Sensitivity

As in the case of LEDs, also OLEDs perform better at low temperatures (efficacy and lifetime) than at high temperatures. The modeling of the stacked structure and the use of temperature-tolerant material are critical factors in obtaining acceptable ambient-temperature sensitivities. Relevant information on the actual temperature sensitivity of a specific OLED type has to be obtained from the manufacturer.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Phosphors and Fluorescent Powders](#)

Acknowledgments Adapted by permission from Springer Nature “[Interior Lighting, fundamentals, technology and application](#)” by Wout van Bommel (2019). Modified and reproduced from Ref. [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

References

1. Van Bommel WJM, Rouhana A. Lighting Hardware: Lamps, Gear, Luminares, Controls. Coursebook, Philips Lighting, Eindhoven (2012)
2. Schubert, E.F.: Light Emitting Diodes, 2nd edn. Cambridge University Press, Cambridge (2006) CrossRef
3. Mottier, P.: Led for Lighting Applications. Wiley, Hoboken (2010)

4. DiLaura DL, Houser K, Mistrick R, Steffy G. Illuminating Engineering Society of North America, New York. 10th (2011)
5. DOE (2017) Solid-State Lighting 2017 Suggested Research Topics. U.S. Department of Energy, office of energy efficiency & renewable energy, Washington

Light-in-Flight Imaging

► Transient Imaging

Lighting Arc

► Light Stage

Lighting Control Systems

► Lighting Controls

Lighting Controls

Peter Dehoff
Zumtobel Lighting, Dornbirn, Austria

Synonyms

[Lighting control systems](#); [Lighting management](#)

Definition

Lighting controls are intelligent network-based lighting solutions. They incorporate the communication between various system inputs and outputs related to lighting control. Input can be given manually by humans and/or by sensor signals or parameter settings. Part of the control systems are unusually one or more local or central computing devices.

Basics

Lighting control is understood to mean the manual or automatic tuning of a lighting situation inside a building, fixed to a building, or in an outdoor area. In its simplest form, the tuning is performed using a light switch or dimmer; in more complex configurations, sensors are used as well as control curves (time lines) and/or control inputs. As a basic requirement, the lighting installation must allow the light sources it incorporates to be varied individually, in groups, or all together. As a minimum, this means the ability to alter the brightness of the light sources but can also include light color and/or spectral composition or even colors. The lighting installation incorporates not only the light fittings (luminaires), operator control units, and control devices but also the blinds or other shading devices which regulate the amount of daylight entering the building. The emergency lighting system can also be integrated. In the case of building management systems, the lighting installation is linked up with other automated services such as heating, ventilation, and air-conditioning.

This type of integrated lighting control is often referred to as lighting management. The purpose of lighting management is to provide the right light in the right place at the right time [1–3].

Overview

The usefulness of lighting control is easy to understand as tuning the light for a particular purpose is not an invention of modern times. Human beings intuitively seek the ideal lighting situation. They spend their lives alternating between periods of activity and relaxation in keeping with the rhythm of day and night. Over the course of the day and the year, they seek and experience the change and variation of natural light.

The invention of electric light brought with it the advantage of having “brightness at the push of a switch,” available at all times. However, there was also an inherent technical limitation: the ability to tune the electric light was initially restricted

to switching it on and off. Lighting was a static arrangement. Despite the fact that the incandescent lamp was dimmable from the outset, this feature proved difficult in the case of discharge lamps. The first widely used fluorescent lamps as well as high-intensity discharge lamps proved reluctant when it came to dimming. Both the physics of these lamps and the ballasts required to operate them made it difficult to regulate their light output without considerable effort. Merely switching them on and off therefore became the common practice. Switching different groups of luminaires was the only approach that came anywhere near “lighting control.”

In the meantime, the available technology has vastly improved. Today’s fluorescent lamps can be cost-effectively dimmed using electronic ballasts, and in the case of LED light sources, tuning the brightness was never a problem. High-intensity discharge lamps, on the other hand, continue to hold out against straightforward dimming.

Lighting control is attracting a lot of attention and is considered a necessary component of a lighting solution.

The main reasons for using lighting control are as follows:

- *To increase energy efficiency:* Dimming and switching reduce the amount of energy consumed by the artificial lighting. When users want less or no light, the artificial lighting can be dimmed or switched off. And if sufficient daylight is available, this can partially or entirely replace the artificial lighting.
- *To improve the light quality:* The ability to adjust the lighting environment to suit the activity or individual preference ensures optimum light quality. Light can be used to set the scene in interiors and in outdoor areas.
- *To enhance safety:* Emergency lighting can be integrated.
- *To ensure flexibility:* By readdressing the individual luminaires, it is possible to provide a fast and flexible response to changes in organization and work layouts. Individual lighting scenes (moods) can also be programmed and changed.

Applications

Where should lighting control be used? The emphasis will vary depending on the application:

Shops

Intuitive control points invite users to adapt the lighting scenes to suit the activity – ranging from light for working at the cash desk to attractive color changes in the lounge.

Growing awareness of the need to save energy has paved the way for dimmable luminaires in retail spaces. It is possible to enhance the visual impression of merchandise and architecture by coordinating changes in color temperature and luminance. High comfort, great flexibility, and low maintenance are the characteristic features of these controlled lighting solutions. The light spectrum, for instance, can be optimally adjusted to suit the illuminated object without the complication of having to change filters. Using control points, timelines, or daylight-linked management of the artificial lighting, the general illumination can also be gradually changed in line with expected light levels over the course of the day. The staging of merchandise to create strong emotional appeal and surprise effects is achieved with the appropriate static and dynamic lighting scenes. Making use of available daylight not only saves energy but also helps to make zones within an interior particularly attractive.

Façades

The use of LED technology in conjunction with lighting control systems has brought about a revolution in façade illumination. The subtlest of messages can be communicated using media façades. Dynamic façade design in particular is aimed at directing the gaze and conveying information. While ecological discussion focuses primarily on the amount of scattered light in the night sky, avoiding the unnecessary use of light by defining sensible operating hours would seem more important. The ultimate purpose of staging façades with light is to attract attention. It therefore makes sense for façades to be lit up exclusively in the evening and morning hours when greater numbers of people are circulating. This approach enables the

identity of companies and communities to be highlighted, structures outdoor spaces at night, and assists nighttime perception while simultaneously addressing ecological objectives.

Museums

Lighting control ensures that sensitive exhibits are only exposed to light which is absolutely necessary, i.e., the level of luminance or the light color required for good perception or when visitors are present, e.g., using occupancy sensing. On and off times can be defined for specific hours of the day. Blinds control and daylight sensors ensure that exactly the right amount of daylight is admitted to achieve a balance between the needs of architecture, human well-being, preventing harm to exhibits, and saving energy. The emergency lighting is discreetly and safely integrated into the lighting management system which provides central monitoring and ensures reliable visibility in an emergency situation.

Hotels

In hotel rooms, intuitive control points which allow the guest to make selections according to individual preference are the first priority. The guest sets the lighting environment. With convenient control of the blinds, the levels of artificial light and daylight are adjusted to suit different activities and visual needs such as watching TV, putting on makeup, or reading. Dynamic lighting moods, with flexible definition based on timelines, or controlled according to weather situation or time of the day, significantly affect the well-being of the guests, particularly in hospitality or wellness areas. In entrance zones, lighting moods based on the outdoor light can also optimize visual adaptation for arriving guests, giving them security and orientation. Defined lighting scenes in conference areas support the use of different media and enable the appropriate light to be provided at the press of a button.

Health and Care

A homely atmosphere and care activities call for entirely different lighting scenes which are available at the press of a button. Intuitive operating features to suit the age and physical abilities of the

occupants as well as the accessibility of the control units are the key to successful lighting solutions in hospitals and care homes. The amount of light required by the eye increases with age. For precise visual tasks, the artificial lighting can be individually adjusted by care staff and patients. In addition, the aging eye filters out blue light, affecting biological processes such as the internal clock and cycles of rest and repair. This must be compensated at specific times of the day by spending time outdoors or through biologically effective artificial light with high intensity and a high blue light component. The use of timelines in lighting management systems permits this interplay of artificial light and daylight at appropriate times of the day.

Office

Lighting control optimally addresses individual lighting needs depending on the age and visual task. Concentrated work and increasing communication call for entirely different requirements to be met by the lighting concept. Biologically effective artificial lighting components at the right time of day in addition to daylight help the internal clock and raise alertness. Lighting management systems with a high level of automation can achieve maximum energy savings as well as flexibility in the case of office moves, thanks to time management, daylight-linked control, and occupancy sensing. Employees readily accept the technology when it gives them the freedom to alter their lighting environment. This means providing adequate means of control and small groups of luminaires with assigned access.

Industry

Industry offers particularly high potential for saving energy. Long periods of lighting use due to shift and night work as well as a lack of daylight in many work areas and rooms mean that investments in lighting management have short payback times. Lighting management ensures the required flexibility in production areas. New lighting installations have to be overdimensioned in order to take into account the effects of deterioration of the installation over time. This additional energy consumption can be counteracted with

daylight-linked or constant light control which continuously adjusts lamp output in line with the available daylight or length of service. Integrated lighting solutions allow the interaction between different services. Maintenance and monitoring operations are also optimized through the incorporation of emergency lighting. Interfaces to other services ensure the convenient and cost-effective operation of buildings.

Education

New forms of teaching and new media technologies call for flexible room use and frequent adaptation of the lighting situation. Intuitive control units are used to select defined lighting scenes at the press of a button, i.e., for working in small groups or traditional teaching, reduced lighting levels for presentations using LCD projectors, or higher vertical illuminance for blackboards and flip charts. Comfortable control units enable the user to rapidly adapt the lighting situation.

Daylight activates human beings and enhances their feeling of well-being as well as their performance. With daylight-linked control of the artificial lighting or occupancy sensing, it is possible to achieve maximum energy savings without restricting the quality of light. Blinds control not only improves the contrasts of presentation media but also increases room comfort because glare and heat gain can be minimized. Conveniently placed control units near the door and on the teacher's desk enable intuitive selection of the appropriate lighting situation.

Corridors and Stairways

Circulation zones are predestined for lighting management. Ideally, the lighting should only be activated if a person approaches. The light should be on once the person has entered the circulation zone. Movement detectors must be suitably positioned. The energy consumption of lighting in corridors and stairways can be further reduced in conjunction with daylight-linked control.

Streets and Public Spaces

Illumination is necessary for reasons of safety and orientation in streets and public spaces. Switching the lighting on and off in conjunction with the

available daylight has long been a common practice. Nowadays, it is also possible to reduce brightness levels at hours of the night when traffic density decreases. Control is performed centrally for streets, districts, and public spaces.

Movement detectors can be installed to ensure that the brightness of the lighting anticipates the approach of pedestrians and cyclists. While the general brightness level is lowered, it can be increased at the point of use to ensure safe circulation.

Domestic

Being able to vary the lighting takes on particular emotional importance in the home setting. Well-being and work, social interaction, and intimacy all take place within the confines of a small space. Access to different lighting moods at the press of a button enables the occupants to set the scene for their private space with area, accent, and mood lighting. Blinds are frequently integrated into the control system to regulate solar heat gain. Time control is also a security feature which can be used to make the house appear occupied when the residents are away on holiday.

Individual lighting scenes can be selected from a central control panel. Switches incorporating a reduced number of scenes on doors and in readily accessible locations allow intuitive operation of the lighting installation.

The possibilities and the requirements of the application are taken into account in the lighting concept. The lighting designer incorporates the operator control units, the control devices, and the choice, number, and arrangement of luminaires, blinds, and other actuators.

Interfaces for Lighting Management

Generally, there are interfaces between the user, the environment, and the lighting installation.

The user can actively intervene using different operator control units:

- *Standard or momentary action switch, with or without dimming function:* The simplest form is a switch and/or dimmer frequently fitted near

the door or in another readily accessible location in the room.

- *Multifunction switch*: Also near the door or in a readily accessible location, for calling up different lighting scenes.
- *Remote control*: This can be an IR or wireless unit or the user's mobile phone, for calling up individual lighting scenes.
- *Workstation control*: In this case, operation is through the user's own computer or using a control unit directly at the workstation.
- *Central control unit*: At a defined location, trained personnel can select specific scenes or timelines for the lighting installation, often using a touch screen.

The environment can be integrated by means of sensors or defined timelines:

- *Daylight sensors*: Sensors installed outdoors or near daylight openings detect the level of daylight brightness and transmit the appropriate signals to control or regulate the lighting installation (luminaires and blinds).
- *Presence or absence detectors*: These sense whether there is a person within their detection zone. The signals they transmit switch the lighting installation on or off.
- *Time signals*: The lighting can be switched to preselected scenes at set times. The simplest case is to switch off all the lights at specific times. It is also possible to call up timelines to automatically tune the lighting in accordance with programmed strategies.
- *User behavior sensors*: Signals based not only on presence or absence but also on user movement patterns can be used to initiate situation-based tuning of the lighting situation.
- *Environmental signals*: Other predefined data can be used for control purposes, e.g., to maintain constant luminous flux on the basis of set maintenance factor curves or to reduce installation output if the energy demand exceeds a specific limit.

Behind each operator function, there are signals which are processed by the lighting management system. This means, for instance, that if a

sensor detects that a person has left the room, switch off can be delayed.

Intelligent control is based on the following logic: for each activity, a lighting scene is defined, saved, called up, and modified as required. Each scene is assigned a name and a symbol which are shown on the operator control units. The lighting scenes are activated automatically by means of presence detectors or simple time inputs. The basic philosophy is always the same: it must be possible to manually override any activated lighting scene.

Strategies for Lighting Management

What strategies can be used to achieve useful objectives with a lighting management system? Firstly, those objectives have to be defined, and, secondly, acceptance of the usefulness of the objectives put to the test.

One objective is to increase energy efficiency:

1. Individual switch on and off: This simplest objective is dependent on a switch being located near the user. Users appreciate this possibility but frequently do not use it.
2. Daylight-linked control – switching function: Automatic switch on and off; the control input is the daylight measured by a sensor.
3. Daylight-linked control – switching function: Automatic switch off and manual/individual switch on; this requires a daylight sensor and a switch.
4. Daylight-linked control – dimming: Automatic setting at a constant value; the control input is the daylight measured by a sensor.
5. Daylight-linked control – dimming: Automatic switch off and individual switch on; this requires a daylight sensor and a switch.
6. Occupancy-based control – switching function: Automatic switch on and off; the control input is delivered by a sensor which detects human presence.
7. Occupancy-based control – switching function: Automatic switch off and individual switch on; this requires a presence detector and a switch.

8. Constant light control (without daylight): A constant level of luminous flux is maintained to compensate for the decrease in light output with increasing age of the installation; the control input is delivered by a programmed timeline or a sensor for the emitted luminous flux.
9. Time-based control: Programmed lighting scenes are called up, and the installation is switched on or off at set times.
10. Load shedding: Automatic limitation of energy demand.

Another objective is to improve the quality of light:

11. Activity-related control: Individual setting of a lighting scene to suit an activity; this includes the switching and dimming of individual blinds, luminaires, and groups of luminaires in the adjacent area.
12. Individual daylight control: Individual operation of blinds to avoid disturbances such as glare and heat.
13. Automatic daylight control: Automatic control of blinds to reduce heat gain and glare from direct sunlight and excessive outdoor brightness; this requires a daylight sensor.
14. Algorithmic lighting: Automatic sequence of variations in the lighting based on programmed rules.
15. Room-related scene setting: Selection of pre-set static or dynamic lighting scenes.

User Acceptance and Energy Efficiency

While the various strategies might bring great benefits from a planning perspective, they also have to be accepted by the users. Many field studies have looked at the question of user acceptance.

One of the essential requirements is the ability to manually override each activated lighting scene. In comparison with automated processes, however, this makes it difficult to calculate the parameters for possible energy savings, the determination of which is a major challenge in view of the many unknown factors.

The behavior of energy-conscious occupants receives too little attention in calculations of this kind. Making the groups of luminaires as small as possible provides greater opportunity to change the lighting situation manually and individually to suit requirements. Large rooms with several groups of luminaires do not lend themselves to this approach as it is difficult for large numbers of users to agree the preferred setting of the lighting installation. For this reason, a smaller group of luminaires and operator control units must be installed in close proximity to enable the user to make spontaneous changes.

Changing weather conditions and unforeseeable user behavior make it difficult to predict energy usage. The time delay for presence detectors also plays a role.

In view of their complexity, the dynamic changes taking place in the world of work and in building use are continually being analyzed and measured in field studies carried out by different industries. In the context of a lighting management system, human behavior will always be more difficult to predict than technically defined parameters (Table 1).

Technical Requirements

The transmission of information from the operator control units and control devices to the actuators is at the heart of any lighting management system. Actuators are luminaires, blinds, or other equipment integrated into the lighting management system.

Information transmission means the sending and receiving of signals, which can be achieved using various means:

- The signal “on/off/dim” is sent via the power line.
- Signal transmission between operator control units and control devices takes place using the power line.
- Signal transmission between operator control units and control devices and to the powered luminaire/actuator uses a separate control line with standardized control signals (DALI, DMX)

Lighting Controls, Table 1 The table shows the effect of individual strategies on energy efficiency and light quality and assesses the associated user acceptance

Strategy	Operation	Energy efficiency	Light quality	Acceptance
1. Individual		Low	High	High
2. Daylight-linked control – switching	Auto off, auto on	Medium	Medium	Low
3. Daylight-linked control – switching	Auto off, manual on	High	Medium	Medium
4. Daylight-linked control – dimming	Auto off, auto on	High	Medium	Medium
5. Daylight-linked control – dimming	Auto off, manual on	Very high	Medium	High
6. Occupancy-based control – switching	Auto off, auto on	High	Medium	Medium
7. Occupancy-based control – switching	Auto off, manual on	Very high	Medium	High
8. Constant light	Auto	High	Medium	Medium
9. Time-based control	Auto	Very high	Medium	High
10. Load shedding	Auto	High	Medium	Medium
11. Activity-related control	Manual	Medium	Very high	Very high
12. Individual daylight control	Manual	Medium	High	Very high
13. Automatic daylight control	Auto	High	High	High
14. Dynamic lighting	Auto	High	Very high	Very high
15. Scene setting	Manual	Medium	Very high	Very high

Auto automatic

- Operator control units and control devices (e.g., computers) are networked via data lines. Communication is by means of bus (LON, KNX, Luxmate, etc.) or TCP/IP protocols.

Definitions

Bus protocols such as LON, KNX, and Luxmate: Bus systems network all technical services in the building and provide central control for heating, ventilation, window blinds, and security systems.

TCP/IP: Transmission Control Protocol/Internet Protocol (TCP/IP) is the set of communication protocols used for the Internet and is consequently also referred to as the Internet protocol suite.

DALI stands for Digital Addressable Lighting Interface and is a standardized digital interface for control gear and electronic ballasts. DALI can be used with a small number of lines to address large numbers of control circuits over large distances.

DMX is an acronym for Digital Multiplex, a standard for digital communication networks commonly used to control stage and event lighting. Nowadays, the DMX protocol is also frequently used by architects and lighting designers as it enables over 500 channels to be controlled individually with rapid signal sequences from a single central control unit.

Luminaires, blinds, and other actuators integrated into the lighting management system have to be compatible with the existing communication protocols and understand the signals. The actuators are assigned individual or group addresses so that they can be uniquely identified within the network and to enable them to communicate.

In the case of classic lighting installations equipped with fluorescent lamps, dimmable ballasts have to be used. High-intensity discharge lamps, on the other hand, can rarely be dimmed, at best switched in stages. As a rule, LED lighting incorporates dimming. With relatively simple means, it also offers the opportunity to vary the light color (with a range from 2,700 to over 8,000 K) or even to set different colors using RGB diodes.

Lighting Design

The decision regarding a lighting management system should be made at an early stage of the project. Lighting management is an integral part of the overall building management system. For this reason, the lighting management system will

often come under a different budget to the actual lighting installation itself.

The choice and layout of the luminaires should allow different lighting scenes to be selected.

The location of operator control units, the layout of the luminaires, and in particular the integration of daylight should be part of a holistic design approach. Service and long-term support for the lighting management system will ensure that the operators obtain the expected benefits.

Factors Driving Lighting Management

The growing demands for lighting installations to be energy efficient are currently a strong driver for the use of lighting management. European green building directives are leading to national regulations for energy requirements in buildings, which cannot be met without the use of lighting management. Daylight-linked control and presence detectors are therefore a necessity.

Green building certification schemes also look to reduce environmental impact with the aid of intelligent lighting management. Well-known schemes include LEED, BREEAM, DGNB, and CASBEE, among others. As well as helping to reduce energy consumption, the integration of lighting management can improve the quality of light for the users.

Cross-References

- [Architectural Lighting](#)
- [Electrical Control Gear for Lamps](#)
- [Interior Lighting](#)
- [Road Lighting](#)

References

1. Craig, D.L.: Lighting controls handbook. The Fairmount Press, Lilburn (2007)
2. Simpson, R.S.: Lighting control. Focal Press, Oxford (2003)
3. David, D.L., Houser, K.W., Mistrich, R.G., Steffy, G. R.: The lighting handbook, 10th edn. IES, New York (2011)

Lighting Design

- [Architectural Lighting](#)

Lighting Dome

- [Light Stage](#)

Lighting for Indoor Spaces

- [Interior Lighting](#)

Lighting for Interior Spaces

- [Interior Lighting](#)

Lighting for Visual Comfort

- [Lighting for Visual Satisfaction](#)

Lighting for Visual Satisfaction

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Lighting for visual comfort](#)

Definition

Lighting that makes the overall appearance of a lit interior space being experienced as visually

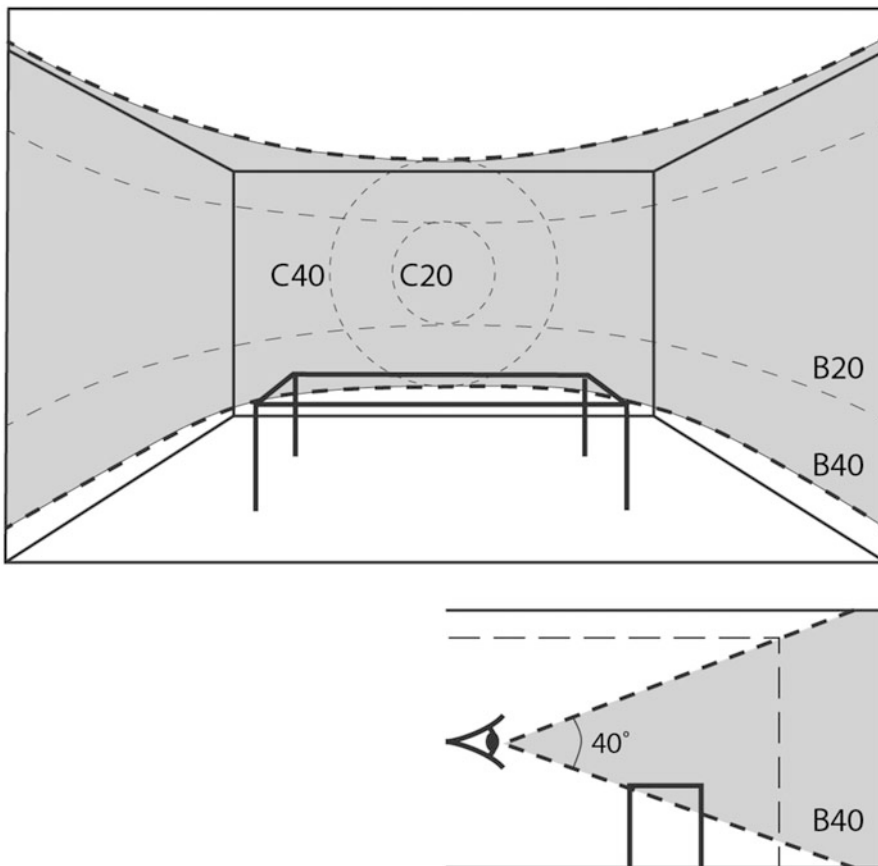
satisfying. Here visual satisfaction relates to such aspects as spaciousness, attractiveness, relaxedness, intimacy, and worker's stimulation.

Room Appearance

The visual appearance of a room determines for an important part whether the lighting has a positive impact on persons in that room. Most of the psychophysical studies that have demonstrated this relate to office type of spaces. These studies try to relate subjective aspects with the luminance and luminance distribution of the surfaces in the space. For this purpose test persons assess a large number of different lighting situations, usually in test rooms under laboratory conditions. The

lighting of the room boundaries determines, to a large extent, the room appearance as a whole. The classic assessment study of room boundary surfaces is that of Loe et al. [1]. The full-scale mock-up office room used for the experiments (Fig. 1) was furnished as a small conference room.

Eighteen widely different lighting situations were assessed by observers from a fixed position in the middle of the long side of the room. Lighting varied in both lighting level and lighting distribution. Detailed luminance measurements of all room surfaces were made for all lighting situations. From these, the average luminance, L_{av} , and luminance uniformity expressed as L_{max}/L_{min} were calculated for the whole area, for the areas represented by circles subtending an angle of 20° and 40° , respectively, and for horizontal bands

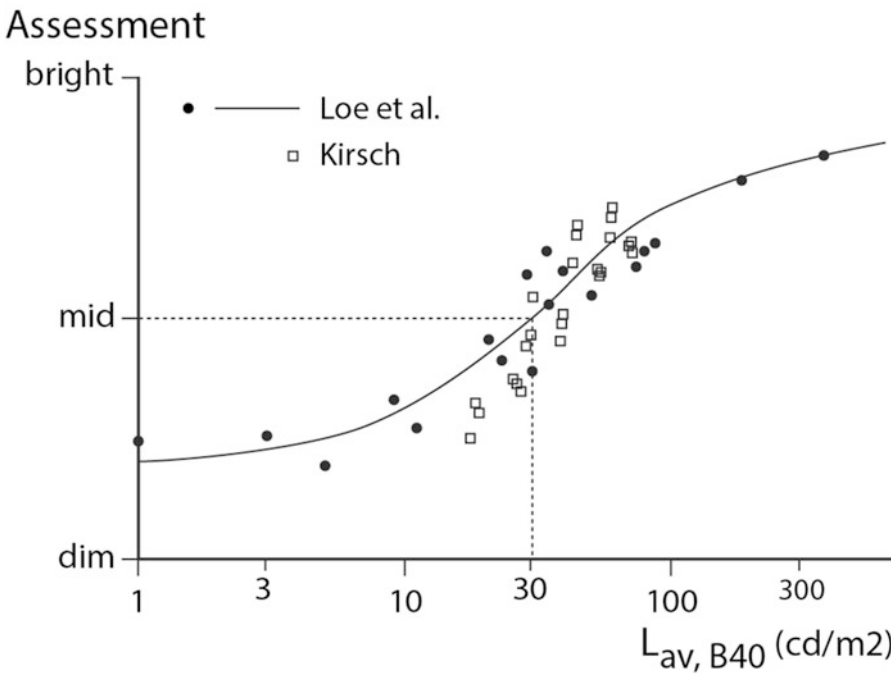


Lighting for Visual Satisfaction, Fig. 1 Mock-up conference room with different dominant areas that were investigated by Loe et al. (1994)

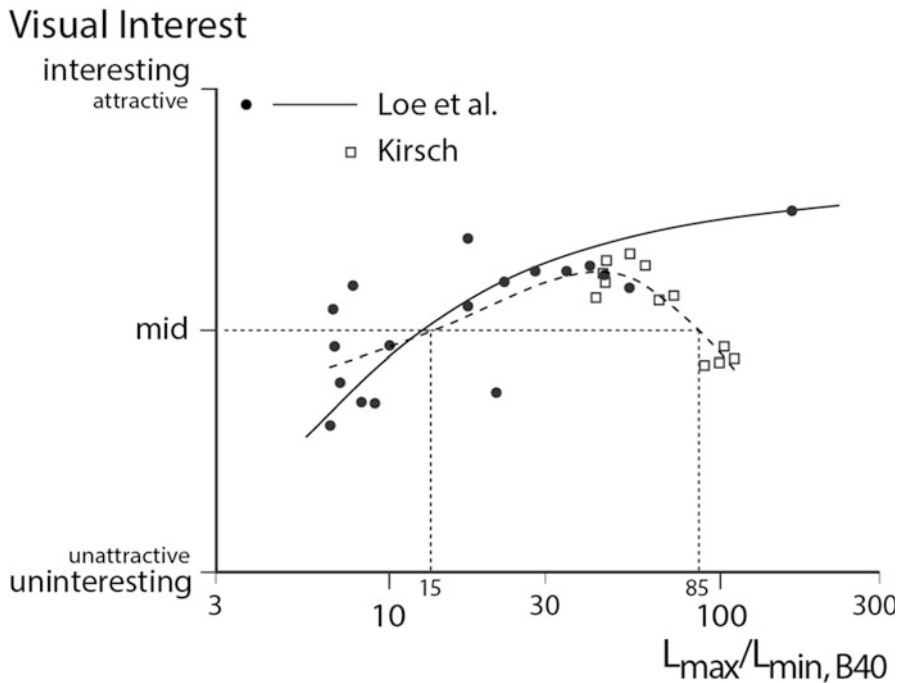
with a width of 20° and 40°, as illustrated in Fig. 1. The 20° circle and band normally only include wall surfaces, while the 40° circle and band, for larger rooms, also include part of the floor and ceiling surface. This becomes clear by comparing the small room, sketched with a broken line at the bottom of Fig. 1, with the larger room drawn in a full line. The subjective aspects assessed can be brought together under two factors. The first factor, “visual lightness,” covers the interrelated subjective aspects of “bright–dim,” “radiant–gloomy,” and “spacious–cramped.” The other factor, which can be labeled as “visual interest,” includes the aspects “nonuniform–uniform,” “interesting–uninteresting,” “pleasant–unpleasant,” “stimulating–subdued,” and “dramatic–diffuse.” The factor “visual lightness” appeared to correlate best with the logarithmic value of the average luminance of the band with a width of 40°, i.e., the B40 band (Fig. 2). The factor “visual interest” correlates best with the logarithm of the uniformity ratio $L_{\max}/L_{\min}$, again of the B40 band (Fig. 3).

Twenty years later, Kirsch [2] made a visual lightness assessment study in a cell type of office mock-up. The walls and ceiling consisted of back-lit (3000 K) diffuse acrylic panels that could be dimmed separately to generate many different room surface luminances. The small squares in Fig. 2 give the results. They are in good agreement with the earlier results of Loe et al. Kirsch also checked the chronotype of his test persons. He showed that there is a significant difference between morning and evening chronotypes. Morning types require in the evening higher luminances than in the morning for the same assessment of brightness. With evening types it is just the other way around: they need in the morning higher luminances than in the evening. At the midpoint, the difference is around 20 cd/m^2 .

The small squares in Fig. 3 give the results of Kirsch for luminance uniformity. The thin broken-line curve in Fig. 3 is the best trend curve when the results of Loe et al. and Kirsch are combined. The relationship that this curve suggests makes sense. Uniform room luminance distributions lead to an



Lighting for Visual Satisfaction, Fig. 2 Bright–dim assessment, being representative for the factor “visual lightness,” as a function of the average luminance $L_{\text{av},\text{B40}}$ of the B40 band [1, 2]



Lighting for Visual Satisfaction, Fig. 3 Visual interest as a function of the ratio $L_{\max}/L_{\min}, B40$ of the B40 band [1]. The results for the appraisal attractiveness [2] are given on the same assessment scale

appraisal “uninteresting” and “unattractive.” The more nonuniform the situation becomes (larger $L_{\max}/L_{\min}$ ratio), the more interesting and attractive lighting is appraised until the maximum appraisal is obtained. From that situation onward, the appraisal decreases with a further increase in nonuniformity until it becomes unacceptable.

By calculating the average luminance and luminance ratio for the B40 band, it is possible to evaluate lighting designs with the aid of Figs. 2 and 3, for each specific observer position and each viewing direction.

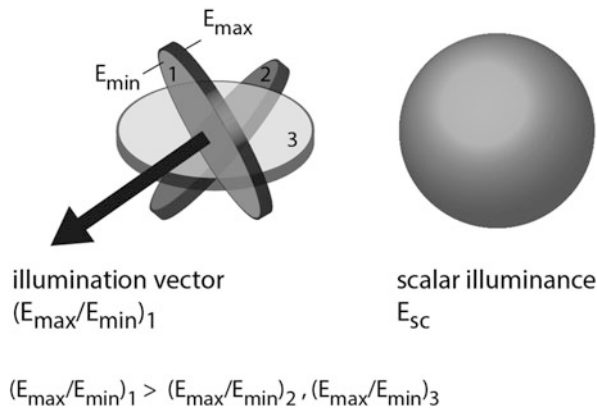
Directionality and Modeling

The directionality of lighting relates to the difference in intensity with which the light is incident on objects from different directions. It determines the appearance of three-dimensional objects and faces of persons in space. A too low degree of directionality (and thus a high diffuseness) gives soft or no shadows, which results in a dull environment in which identification of objects and faces is difficult. Too strong directionality produces harsh shadows

in which details of the object are hidden, making the object appear unattractive and difficult to identify. The modeling effect of lighting is good when it reveals the details and texture of objects and results in an aesthetically pleasing environment. The balance between the diffuse and directional components of a lighting installation determines the quality of directionality.

Flow of Lighting

To mathematically study the effect of light on objects in space, Gershun [3] introduced in 1936 the expression “light field” and visualized it with “flow lines.” Cuttle et al. [4] introduced a promising and practical concept, which they called “flow of lighting.” This concept allows for the calculation of the main direction and strength of the light at a point in space as a result of all light rays at that point. Cuttle showed that flow of lighting could be quantified by the so-called vector-to-scalar ratio, $E_{\text{vector}}/E_{\text{scalar}}$ [4–6]. It contains two measures, the illuminance vector and scalar illuminance. The illuminance vector characterizes the direction and magnitude of the light flow at a point in the lit space. The final directional



Lighting for Visual Satisfaction, Fig. 4 Left: The illumination vector at a point is defined by that (infinitesimally small) disk having the largest difference in illuminances on the opposing sides of the disk. The vector is directed away

from the side having the lowest illuminance. Right: The scalar illuminance is the mean illuminance on the surface of an (infinitesimally small) sphere

effect of it depends on the ambient or overall lighting at that same point. The higher the overall lighting at a point is, the lower the directional effect. The scalar illuminance (E_{scalar}) at a point characterizes the overall illumination at that point. It is the mean illuminance on the surface of an (infinitesimally small) sphere at that point (Fig. 4, right). It is also called mean spherical illuminance. The maximum difference in illuminance between opposite sides of an (infinitesimally small) flat disk at a point defines the magnitude of the illuminance vector (E_{vector}) at that point (Fig. 4, left). The direction of the vector is defined by the orientation of the disk for which the maximum difference is obtained. It is the direction of the normal to the disk surface at that orientation, away from the side having the lowest illuminance (Fig. 4, left).

With no directionality of lighting, the vector/scalar ratio is zero, and the maximum value is 4, which is the case in a situation with a single parallel light beam in a complete black interior. Table 1 gives vector/scalar ratio values with corresponding appearances and suggested applications as proposed by Cuttle.

Light Tubes

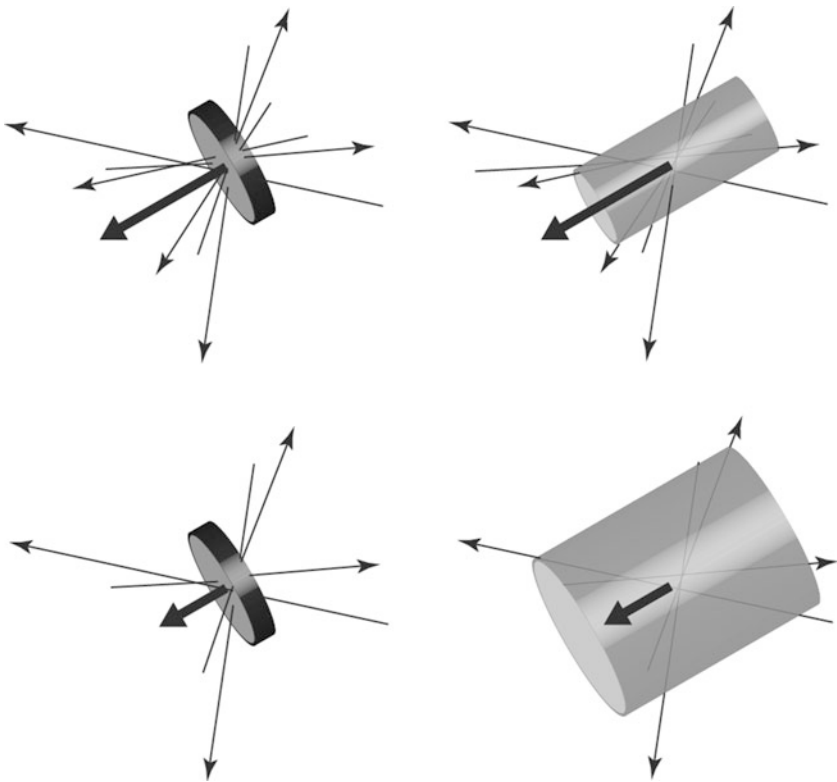
The flow of lighting can be visualized by so-called light tubes [7]. Light tubes are a further

development of the earlier mentioned flow lines of Gershun. The direction of a light tube is in each position parallel to the light vector, while the opening width of the tube is inversely related to the magnitude of that vector. This magnitude is equal to the average light flux “flowing” through the tube at that position. It means that at locations where the tube is small and the magnitude of the vector is large, the net light transport is large (high flux density). Near to light sources, from where the tubes originate, they are small (large flux transport). The tubes end up on the room surfaces, where the net light transport has become small, and the tubes consequently are wide. Light tubes are not the same as light rays and can, therefore, contrary to light rays, be curved. Light tubes are based on the combined effect of all light incident at a point from all directions (from both direct and indirect contributions). Figure 5 illustrates the difference between light rays and light tubes.

With today’s available computer graphic software, methods for light tube visualizations of complete lighting installations have been developed [7–12]. This type of visualizations allows a detailed analysis of the spatial and form-giving potential of lighting designs. It combines two early concepts, “lighting field” (Gershun [3]) and “flow of lighting” (Cuttle et al. [4]), into a modern tool for the professional lighting designer.

Lighting for Visual Satisfaction, Table 1 Vector/scalar ratio with appearance effect and application suggestions (Cuttle [6])

$E_{\text{vector}}/E_{\text{scalar}}$	Appearance	Application
4.0		
3.5	Dramatic	Strong contrasts, details in shadows not discernible
3.0	Very strong	Suitable for displays, too harsh for human features
2.5	Strong	Pleasant appearance for distant faces (formal)
2.0	Moderately strong	Pleasant appearance for near faces (informal)
1.5	Moderately weak	Soft lighting for subdued effects
1.0	Weak	Flat shadow-free lighting
0.5	Very weak	
0		



Lighting for Visual Satisfaction, Fig. 5 Difference between light rays and light flow. Small arrows represent light rays and the broad arrow the illumination vector. The diameter of the light tubes is inversely proportional to the

magnitude of the vector. Top, large illumination vector and corresponding small light tube; bottom, smaller illumination vector and corresponding wider light tube

Discomfort Glare

Glare can take either of the two forms: disability glare and discomfort glare. Disability glare is the form that is responsible for the negative influence of glare on visual performance. Disability glare

has a neglectable effect on visual performance under most interior lighting conditions [13]. Glare that causes a feeling of discomfort is referred to as discomfort glare. It can cause problems in interior lighting situations. The discomfort effect of glare may lead to feelings of irritation and

fatigue. The eyes may contract. In extreme circumstances, discomfort glare may result in painful feelings and after prolonged exposure in headaches. The adverse effects of discomfort glare vary to a great extent in type and seriousness between persons.

The physiological and psychological mechanism underlying discomfort glare sensations is not yet clearly understood. Bargary et al. [14] studied functional magnetic resonance imaging (fMRI) scans of the brain produced in the presence of discomfort glare. Their results suggest that the physiological properties of receptive fields and their ON-OFF center-surround ganglion cells may play a fundamental role in evoking discomfort glare sensations. Donners et al., Scheir et al., and Safdar et al. [15–18] use as a basis for their discomfort glare considerations the neural response to bright light and the mechanism of receptive fields. These studies demonstrate that models built upon center-surround receptive fields and their ganglion cells that transform luminance signals into edge detection signals are promising candidates for quantifying discomfort glare from a physiological point of view. Until a generally accepted fundamental approach, based on the mechanism underlying discomfort glare sensations, results in a workable method to quantify discomfort glare, empirical studies form the background for the metrics used to quantify discomfort glare.

In 1995, CIE combined the best parts of the different, empirically derived, glare evaluation methods in use at that time into one formula. Because of this “consensus method,” the result is called “CIE Unified Glare Rating (UGR)” [19]. This method, recently slightly adapted to make it also suitable for LED luminaires with an irregular light-emitting surface luminance [20], is still in use today for the restriction of discomfort glare caused by artificial lighting in interiors. It is described in more detail in this encyclopedia under the entry “Glare.”

Light Color Preference

Research on preferred correlated color temperature for general lighting in interiors gives conflicting results. Kruithof was the first who did

a study, already in 1941, on the relationship between correlated color temperature and appraisal of the pleasantness of lighting [21]. Up to the present day, his study is often used as a reference. With all the knowledge gathered since Kruithof carried out his tests, the conclusion can only be that his conclusions are not valid for general use. The entry “Kruithof Curve” in this encyclopedia gives a more detailed description of the subject.

Cross-References

- Glare
- Kruithof Curve

References

1. Loe, D.L., Mansfield, K.P., Rowlands, E.: Appearance of lit environment and its relevance in lighting design: Experimental study. *Light. Res. Technol.* **26**(3), 119–133 (1994)
2. Kirsch, R.M.: Lighting quality and energy efficiency in office spaces. Doctoral thesis, Department of Lighting Technology, Technical University, Berlin (2014)
3. Gershun, A.: *Световое поле (light field)*, Moscow. Translated (1939) by Moon PH and Timoshenko G. *J. Math. Phys.* **18**, 51–151 (1936)
4. Cuttle, C., Valentine, W., Lynes, J., Burt, W.: Beyond the working place. In: *Proceedings CIE 16th Session*, pp. 19–28. Washington, DC (1967)
5. Cuttle, C.: Lighting patterns and the flow of light. *Light. Res. Technol.* **3**, 171–189 (1971)
6. Cuttle, C.: *Lighting by Design*, 2nd edn. Architectural Press, Oxford (2008)
7. Mury, A.A., Pont, S.C., Koenderink, J.J.: Representing the light field in finite three-dimensional spaces from sparse discrete samples. *Appl. Opt.* **48**, 450–457 (2009)
8. Pont, S.C.: Spatial and form-giving qualities of light. In: Albertazzi, L. (ed.) *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space and Appearance*. Wiley, Chichester (2013)
9. Huang, A., Sanderson, A.: Light field modelling and interpolation using Kriging techniques. *Light. Res. Technol.* **46**, 219–237 (2014)
10. Kartashova, T., Sekulovski, D., De Ridder, H., Te Pas, S.F., Pont, S.C.: The global structure of the visual light field and its relation to the physical light field. *J. Vis.* **16**(10), 9 (2016)
11. Xia, L., Pont, S.C., Heynderickx, I.: Light diffuseness metric Part 1: Theory. *Light. Res. Technol.* **49**, 411–427 (2017a)
12. Xia, L., Pont, S.C., Heynderickx, I.: Light diffuseness metric Part 2: Describing, measuring and visualising

- the light flow and diffuseness in three-dimensional spaces. *Light. Res. Technol.* **49**, 428–445 (2017b)
13. Van Bommel, W.: *Interior Lighting, Fundamentals, Technology and Application*. Springer, Cham (2019)
 14. Bargary, G., Furlen, M., Raynham, P.J., Barbur, J.L., Smith, A.T.: Cortical hyperexcitability and sensitivity to discomfort glare. *Neuropsychologia*. **69**, 194–200 (2015)
 15. Donners, M.A.H., Vissenberg, M.C.J.M., Geerdinck, L.M., Van Den Broek-Cools, J.H.F., Buddemeijer-Lock, A.: A psychophysical model of discomfort glare in both outdoor and indoor applications. In: *Proceedings 27th CIE Session*, pp. 1602–1611. Manchester (2015)
 16. Scheir, G.H., Donners, M., Geerdinck, L.M., Vissenberg, M.C.J.M., Hanselaer, P., Ryckaert, W.R.: A psychophysical model for visual discomfort based on receptive fields. *Light. Res. Technol.* **50**, 205–217 (2018)
 17. Safdar, M., Luo, M.R., Mughal, M.F., Kuai, S., Yang, Y., Fu, L., Zhu, X.: A neural response-based model to predict discomfort glare from luminance image. *Light. Res. Technol.* **50**, 416–428 (2018)
 18. Scheir, G.H., Hanselaer, P., Ryckaert, W.R.: Pupillary light reflex, receptive field mechanism and correction for retinal position for the assessment of visual discomfort. *Light. Res. Technol.* **51**(2), 291–303 (2019)
 19. CIE: International Commission on Illumination CIE Publication 117:1995, Technical Report, Discomfort Glare in Interior Lighting, Vienna (1995)
 20. CIE: International Commission on Illumination CIE Publication 232:2019, Technical Report, Discomfort Caused by Glare From Luminaires with a Non-uniform Source Luminance, Vienna (2019)
 21. Kruithof, A.A.: Tubular luminescence lamps for general illumination. *Philip. Tech. Rev.* **6–3**, 65–73 (1941)

Lighting Management

► Lighting Controls

Linguistic Influences on Color Perception

► Effects of Color Terms on Color Perception and Cognition

Linguistic Relativism

► Color Category Learning in Naming-Game Simulations

Lippmann, Jonas Ferdinand Gabriel

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



Gabriel Lippmann was a French inventor and physicist (born in Luxembourg) who created the first color photographs using what was certainly the first spectral imaging system. Lippmann's system of color photography was conceived in 1886 and then refined for several years due to the complex nature of its theory and implementation. The system works by placing a very fine-grain photographic plate in contact with mercury that acts as a mirror. Light waves pass through the emulsion, reflect from the mercury backing, and then create an interference pattern within the emulsion. The developed plate then has an interference filter built into it due to the properly spaced layers of silver in the emulsion (created by the interference pattern exposure). The plates are then viewed with directional lighting, and the observer sees the same wavelengths that were present in the scene – a spectral image reproduction. Lippmann presented

the first color photograph using his system in 1891 and then presented several nearly flawless photographs made by Auguste and Louis Lumière created with the process. The process was difficult and time-consuming, and few have been able to replicate the stunning photographs. Lippmann's work certainly presaged modern photographic and holographic processes, and for it he received the 1908 Nobel Prize in Physics.

Despite ending up a full professor at the Sorbonne (University of Paris), Lippmann had no formal education beyond high school. He was a student at the École Normale, but failed the examination that would have qualified him as a teacher due to his penchant for concentrating only on the work that interested him and neglecting the rest. However, he was appointed to a government scientific mission to Germany where he was able to work with the likes of Kirchhoff and Helmholtz. At about the same time, in 1873, he invented the Lippmann capillary electrometer for precise measurements of extremely small electrical voltages. It served as the basis for early echocardiographs. Lippmann joined the Faculty of Science in Paris in 1878, became Professor of Mathematical Physics in 1883, and was later appointed Professor of Experimental Physics and Director of the Research Laboratory. He made many contributions in various fields of physics including electricity, thermodynamics, optics, and photochemistry.

In addition to his Nobel Prize for the Lippmann process of full-color photography, Gabriel Lippmann served as Marie Curie's thesis advisor at the Sorbonne and let her use his laboratory for her thesis work in radioactivity and helped her find other sources of support. Lippmann died at sea in 1921 while returning from a voyage to Canada. There is no record of the cause of death.

Lippmann's Nobel lecture on color photography can be found at the following link:

http://www.nobelprize.org/nobel_prizes/physics/laureates/1908/lippmann-lecture.html

Logic

► [Color Categories and Vagueness](#)

Lorenz-Mie Theory

► [Rayleigh and Mie Scattering](#)

Lorenz-Mie-Debye Theory

► [Rayleigh and Mie Scattering](#)

Lovibond, Joseph Williams

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



Joseph Lovibond was a British chemist and brewer and is credited with inventing the commercial colorimetry, the Lovibond Tintometer. In Lovibond's *Light and Colour Theories* [1], we read "the writer was formerly a brewer, and this work had its origin in an observation that the finest flavour of beer was always associated with a colour technically called 'golden amber,' and that, as the flavour deteriorated, so the colour assumed a reddish hue." Such observations of the relationships between beer quality and observed color led

Lovibond to his work on color standards as a reliable means of reference and to the development of an instrument, the visual colorimeter, in which such standards could be systematically and objectively applied.

The general form of the Lovibond Tintometer was a visual colorimeter in which a split field is viewed. One half of the split field represented the test sample, perhaps a cuvette of beer placed in a beam of light. The other side of the split field was composed of subtractive primaries of adjustable density (e.g., a series of cyan, magenta, yellow, and neutral glass filters) that could be adjusted to select the density and overlapped one another in the adjacent beam of light. By adjusting the density of the filters, observers would match the test color stimulus and record the densities of the standard filters required for the match. Note again that the Tintometer system was one of subtractive color mixing of standard materials rather than the more common additive color mixing of standard lights typically found in visual colorimetry.

While brewing beer might have been the motivation for the Tintometer, by 1914, the system was in use in a wide variety of additional industries including tanning, wine and spirits, dyeing and printing, paint, water chemistry, ceramics, various oils, and hematology. In many cases, specific versions of the instrument and the reference standards were produced for a given application. The system was also presented with a number of awards by international juries (including two gold, five silver, and two bronze medals) along with significant recognition from ten scientific societies (one gold, three silver, and five bronze medals and a diploma). Lovibond worked tirelessly in promoting his system through lectures and demonstrations around the world. He created a technically successful system that also met with commercial success. In fact, The Tintometer Limited still exists to this day with products such as color standards and scales, visual colorimeters (including versions of the Lovibond Tintometer), and photoelectric colorimeters and spectrophotometers. For example, one can still purchase instruments and standard color scales designed

for American (ASBC) and European (EBC) methods of specifying beer color. Lovibond's initial inspiration is still being addressed by the progeny of his instruments using his very techniques.

In the history of The Tintometer Limited, it is stated that "The Company was founded in 1885 by Joseph Lovibond, a prominent brewery owner who developed the 'colorimeter' as a means of ensuring the quality of his beer. By 1893 he had perfected his research and introduced the first instruments. Much has developed since then. Today, the company is bringing colour measurement to the next generation. While still recognizing the importance of traditional methods, The Tintometer Ltd is introducing new techniques to bring measurement and quality control to an even higher level, developing creative solutions to ensure the continued reputation of the Lovibond® brand" [2].

References

1. Lovibond, J.W.: *Light and Colour Theories and Their Relation to Light and Colour Standardization*. E. & F. N Spon, Haymarket, London (1921)
2. www.lovibond.com

Low-Pressure Glow Discharge Lamps

► Neon Lamp

Lumen Depreciation

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Light depreciation](#)

Definition

Light losses during the use of a lighting installation, caused by the decrease of the luminous flux of lamps, the dirt accumulation on and loss of reflectivity and/or transparency of luminaires, and dirt accumulation on or discoloration of room surfaces (in interiors).

Lamp Lumen Depreciation

The light output of virtually all lamp types declines gradually during operation. The causes of light output depreciation are numerous. With incandescent lamps, it is mainly the blackening of the bulb – caused by evaporation of the filament. Discharge lamps also suffer from blackening, in this case, due to scattering of the electrode material, which settles on the wall of the discharge tube. With fluorescent lamps, high-pressure mercury lamps, and solid-state LED and OLED light sources with a fluorescent coating, the primary cause of light output depreciation is a gradual exhaustion of the fluorescent powder, which slowly loses its effectiveness.

Because of the different causes of lamp lumen depreciation, the actual rate of decline of lumen output is different per lamp type. Lamp manufacturers should be able to supply lamp lumen depreciation curves for their lamps.

Lumen Depreciation and Lifetime of LEDs

Quality non-retrofit LED light sources have nowadays such a long life, 50,000 hours and longer, that it is not the failure rate that determines how long the LED light source continues to provide lighting up to specification. It is in particular the LED parameters “lumen output” and “color quality” that determine operational life. When the value of one of these parameters becomes so low that the LED light source has to be replaced, that LED source has a so-called parametric failure, although the LED source may continue to

function for a much longer time. The moment at which the LED light source completely stops functioning is referred to as an abrupt or catastrophic failure.

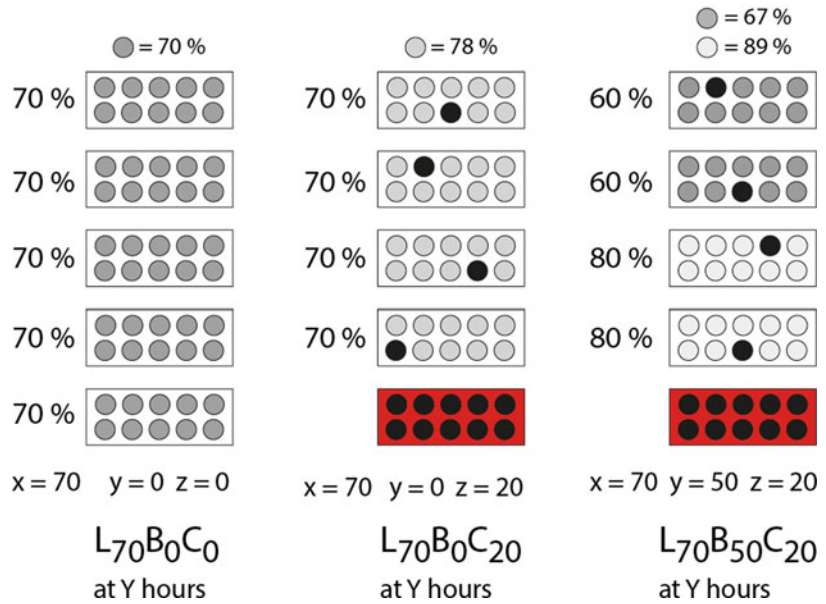
The International Electrotechnical Commission (IEC) defines in its standard “LED modules for general lighting – Performance requirements” [1], LED lifetime by the parametric failure: “too low lumen output.” This lifetime is specified as Y hours useful life based on the condition Lx B50. Here Y is the time (in hours) after which 50% (B50) of a population of LED modules parametrically fails to provide at least a percentage x of the initial luminous flux. Which lumen depreciation percentage is relevant is dependent upon the type of application. L90B50, L80B50, and L70B50 are values that may be relevant. The L70 condition is often used as corresponding to “the lifetime” of a LED source. However, for quite some applications, a lumen maintained value of 70% is too low and requires unneeded high new lighting values to be installed. Just using L70 increases the chances of specifying low-quality products. High-quality products with long lifetime values corresponding to L80 or L90 conditions are often more cost- and energy-effective. It is important to stress that lifetime specifications without indication according to which Lx condition are senseless. According to IEC the value of B always has to be 50, meaning that 50% of a set of LEDs of the same type failed to deliver the declared percentage x of lamp lumen, i.e., B50. If only the Lx value is given, B is assumed to be 50.

An essential point in the IEC standard is that it considers failures of single LEDs in a multi-LED module to contribute to the gradual lumen output degradation. The effect is incorporated in the lamp lumen depreciation and thus in the Lx condition. Abrupt, catastrophic, complete failures of a LED module are referred to as Cz: length of time during which z% of a population of LED modules of the same type fail to produce any luminous flux. Figure 1 gives typical examples of different LxBxCz conditions for a LED module.

An accurate prediction of the percentage of abrupt complete failures Cz over the very long lifetime of LEDs is practically not possible. It is

Lumen Depreciation,

Fig. 1 Examples of different LxByCz conditions of a LED module



also not so important because, as discussed, it is the lamp lumen depreciation, and thus Lx, that determines the useful life of LEDs.

The industry standard for lumen maintenance testing of a batch of LEDs is a document of the Illuminating Engineering Society of North America, “IES Document LM-80, Measuring lumen maintenance of LED light sources” [2, 3]. IEC [1] refers to these documents for the testing of lumen maintenance of LED light sources. The LM-80 document prescribes the LED sampling method, the laboratory environment conditions, the photometric measurement protocol, and the operating conditions (electrical and thermal) of the LEDs to be tested.

Tests according to document LM-80 provide no lifetime of the LED type tested. The data obtained are the input for another IES report that predicts through interpolation of the LM-80 data, useful life. This second IES document, TM-21 “Projecting long term lumen maintenance of LED light sources” [4], describes a mathematical interpolation method, using data obtained with the test described in LM-80. It permits for a prediction of the lumen maintenance percentage at a life six times that of the tested period (based on a minimum sample size of 20 pieces). More extended predictions are unrealistic because of limitations

in the extrapolation and lack of confidence in the data beyond. A more recent addendum of TM-21 stresses again that calculated (interpolated) results beyond the limit of six do not have a practical or meaningful value [5]. For an LM-80 tested period of 6000 h, this corresponds to a prediction for maximum 36,000 h. For a tested period of 10,000 h, the prediction is for maximum 60,000 h.

Luminaire Depreciation

The gradual reduction of output of a luminaire is caused by dirt that is gradually deposited on lamps and on or in the luminaires and by a gradual loss of reflectivity of reflectors and mirrors or loss of transparency of refractors due to corrosion and discoloration. Where the effects of dirt accumulation can be offset by regular cleaning, the output loss due to corrosion and discoloration cannot be regained. The rate of output reduction depends on the materials used in the luminaire, the construction of the luminaire, and the nature of airborne dirt in the atmosphere. The degree of protection provided by the construction of the luminaire is classified according to the international protection code, IP, as described in an international IEC standard.

Room Surface Depreciation

Gradual dirt collection on room surfaces and/or discoloration of these surfaces gradually reduces the interreflected component of light from the installation. This may especially have a noticeable effect in interiors of smaller dimensions where, because of the relative large wall area, interreflection components are relatively large.

Maintenance Factor

The maintenance factor is the ratio of illuminance or luminance produced by the lighting installation at the moment that maintenance operations are being carried out to the illuminance or luminance produced by that installation when new. The maintenance factor takes into account all light losses described above together with the average light loss of the installation due to failed lamps not being replaced. The actual value is, of course, also dependent on the length of the period between maintenance operations. The term light loss factor has sometimes been used instead of maintenance factor. Depreciation factor has been formerly used to designate the reciprocal of the above ratio.

The recommended lighting level (in terms of illuminance or luminance) for a lighting design is based on “maintained lighting level” which is the average lighting level at the moment that maintenance is being carried out. This implies that the initial lighting level has to be:

$$L_{\text{initial}} = L_{\text{maintained}} / \text{Maintenance factor}$$

It is the responsibility of the lighting designer to make an economical balance between the recommended length of the maintenance interval and the extra light required for the initial installation. Shorter maintenance intervals increase the maintenance cost but allow for lower initial lighting levels with a lower initial investment and lower energy costs.

Cross-References

- [Interior Lighting](#)
- [Luminaires](#)

- [Road Lighting](#)
- [Sports Lighting](#)

References

1. IEC: IEC 62717 edition 1.1. LED modules for general lighting – performance requirements (2015)
2. IES: LM-80-08 Approved method: measuring lumen maintenance of LED light sources (2008)
3. IES: Errata #1 for LM-80-08 Approved method: measuring lumen maintenance of LED light sources (2015)
4. IES: TM-21-11 Projecting long term lumen maintenance of LED light sources (2011)
5. IES: Addendum B for TM-21-11 Projecting long term lumen maintenance of LED light sources (2015)

Luminaires

Wout van Bommel

van Bommel Lighting Consultant, Nuenen, The Netherlands

Synonyms

[Fitting](#) (Anglo-Saxon English); [Fixture](#) (American English)

Definition

Devices that control the distribution of the light emitted by a lamp or lamps and that include all the items necessary for fixing and protecting the lamps (and sometimes the gear, too) and for connecting them to the electricity-supply circuit [1].

Luminaire Characteristics

The principal characteristics of luminaires can be listed under the following headings:

- Optical
- Mechanical
- Electrical
- Thermal
- Aesthetics

Optical Characteristics

The optical characteristics of a luminaire determine the shape of its light beam or light-output distribution. The light distribution of a luminaire defines how the luminous flux radiated by the lamp or lamps is distributed in the various directions within the space around it. Different lighting applications require different light distributions and thus different luminaires.

The desired light distribution of a luminaire is obtained through the application of one or more of the physical phenomena: reflection, refraction, and diffuse transmission. Many luminaires also make use of shielding in one form or another, principally to obtain the required degree of glare control and to limit light pollution. The shielding function may be performed by refractors or diffusers or by mirror reflectors, by white-painted surfaces, or, where very stringent glare control is required, by black surfaces.

Reflectors

Many conventional luminaires are provided with a reflector (sometimes in conjunction with another light-control element) in order to create the appropriate light distribution. The reflecting material that is used for reflectors can be specular, spread, or diffuse.

Specular Reflectors

Specular reflectors (also called high-gloss mirror reflectors) are used when a precise form of light distribution is required, as in floodlights, spotlights, and road-lighting luminaires. The reflector creates multiple images of the light source. The most widely used material is sheet aluminum, which has the strength needed to produce a stable reflector. Reflectance values are around 0.70. Alternatively, commercial-grade aluminum can be clad with a thin layer of super-purity aluminum or silver. With aluminum, reflectance values of up to 0.80 can be obtained, while with silver a reflectance of more than 0.90 is possible. Finally, there is vacuum metalizing, in which a specular layer of aluminum is deposited on a suitably smooth

substrate (metal, glass, or plastics). The resulting reflectance, which is somewhere between 0.80 and 0.90, is dependent on both the substrate material and the quality of the metalizing process.

Spread Reflectors

With spread reflectors (sometimes also called half-matt reflectors), there is no sharp mirror image of the light source. They are employed where a moderate degree of optical control is required, with the emphasis on producing a beam with smooth transitions. Such reflectors also help to smooth out discontinuities in the light distribution caused by inaccuracies in the shape of the reflector. Spread reflection is produced by brushing or etching or by hammering very small dimples and bumps into a specular surface (Fig. 1).

Diffuse Reflectors

At the other extreme from specular reflection is the diffuse reflection, which is also called matt reflection. Here light incident on the reflector is scattered in all directions, so there is no mirror image of the light source. Matt reflectors cannot provide sharp beam control but are employed where diffuse or non-focused light distributions are required. Matt-finished metals and glossy white paints on metal or glossy-white plastics provide near-diffuse reflection. The small specular component due to the gloss is of no practical optical significance; the gloss merely serves to facilitate cleaning. Reflectance values can be in the range 0.85–0.90. Ceramic materials or finishes have completely diffuse reflection characteristics with extremely high reflectances of up to 0.98.

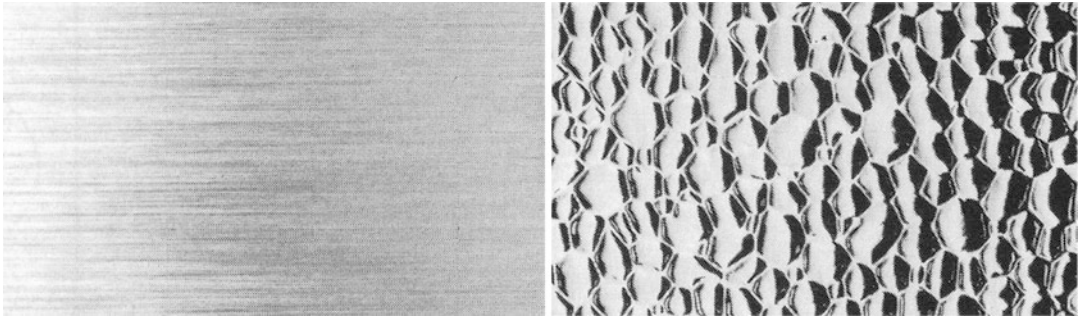
Reflector Forms

There are three basic reflector forms: plane, curved, and faceted (Fig. 2).

Plane Reflectors

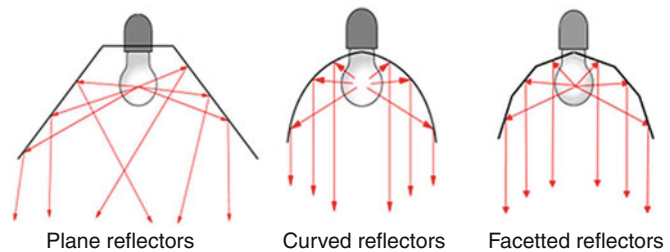
When using a simple plane (or straight-sided) reflector, the light emitted by the light source is reflected according to the material of the reflecting surface, viz., specularly or diffusely.

Plane reflectors are often used to screen off the direct light from the light source. Accurate beam



Luminaire, Fig. 1 A spread finish as produced by brushing (*left*) or by hammering (*right*) a specular surface

Luminaire, Fig. 2 Basic reflector forms



shaping is scarcely possible with plane reflectors, but by changing the symmetry of the reflectors, the direction in which the bulk of the light is emitted can be changed.

Curved Reflectors

The best optical performance is obtained when using a curved reflector. Depending on the curvature, many different types of beams can be created. A curved reflector may be cylindrical, parabolic, elliptical, hyperbolic, or some other contour to suit a particular application. The circular and parabolically shaped reflectors are the ones most commonly used.

The most important optical property of a parabolic reflector is that a point source of light placed in its focus will produce a parallel beam of reflected rays with the highest intensity in its center. If the light source is not at the focus but in front or behind it, the reflected rays are no longer parallel. Thus, by choosing the position of the light source relative to the focus point, the desired beam shape (narrow to wide) can be

created. Since a lamp is never a real point source, deviations from the theoretical beam shape for a point light source as sketched above will always occur. The smaller the light source relative to the size of the reflector, the more accurately can the beam be shaped.

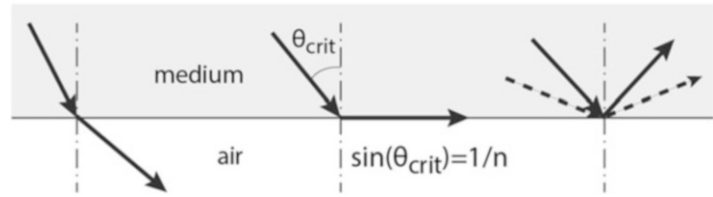
Faceted Reflectors

Smooth-curved reflectors have to be produced to a high degree of accuracy because even small deviations from their intended shape will produce undesirable discontinuities in their light distribution (striations). This will not occur with a faceted reflector. A faceted reflector consists of a number of adjacent, plane or curved, facets that together approximate a curve that is an approximation of a parabolic curve. The width of beam produced by the faceted reflector is somewhat larger than that of a smooth-curved reflector.

Total Internal Reflection Optics

The phenomenon of total internal reflection (TIR) on the boundary between an optical medium and

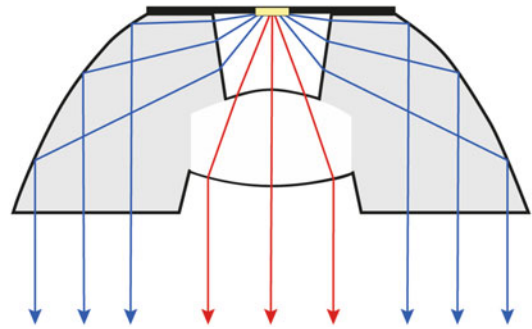
Luminaires, Fig. 3 Refraction and, for $\theta > \theta_{\text{crit}}$, total internal reflection (TIR)



air can also be used to shape the light distribution. Total internal reflection may occur if light travels from a medium with a higher optical density to a medium with a lower optical density, i.e., from high to low refraction index. Glass and plastics have higher refraction index values than air. This means that at the boundary between the medium and air, refraction is away from the normal to the boundary layer, as shown in Fig. 3-left. The middle picture of Fig. 3 shows the situation where the light incidence angle is such that the refraction angle is 90° and thus parallel to the boundary surface. That incidence angle, corresponding to 90° refraction, is called the critical angle. It varies, depending on the type of plastic or glass, between approximately 30° and 40° . If the light incidence is larger than this critical angle, light cannot leave the medium and is reflected with a very high reflection percentage (nearly 100%) as shown in Fig. 3-right.

Total internal reflection with its extremely high reflectance is often used in combination with LEDs to produce efficient and precisely controlled beams in downlight and small floodlight type of luminaires. Often these TIR optics are combined with a collimator lens. Figure 4 shows an example of such a “hybrid” TIR-lens system.

The TIR part controls the largest part of the light radiated from the LED chip, while the integrated collimator lens, located in the center of the system, controls the light radiated from the chip into the central directions. An additional advantage of these TIR-lens systems is that it protects the LED, as it is entirely covered by the bulk optics. For high luminous flux LEDs, bulk plastic optics cannot be used, because of temperature reasons and only glass optics are used. Glass is often also used for lower flux LEDs.



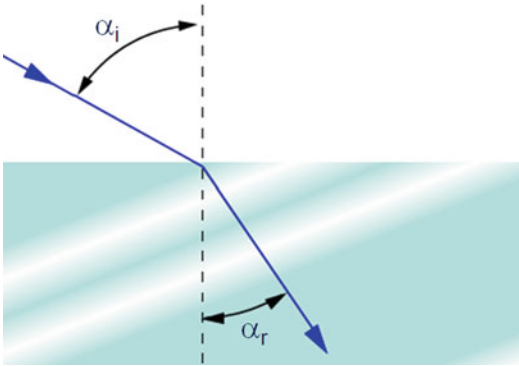
Luminaires, Fig. 4 Example of a rotation-symmetrical TIR-lens solid body for a LED. Blue rays are the result of TIR and red rays the result of the integrated collimator lens (indicated with a lighter tint)

Refractors

Refractors are used to create the desired luminaire light distribution by passing the light from the source through a refractor (Fig. 5). The angle through which the light is bent is dependent on both the shape of the refractive material and its refractive index (Snell's law).

$$\frac{\sin(\alpha_i)}{\sin(\alpha_r)} = \frac{n_{\text{air}}}{n_{\text{material}}}$$

Refracting devices are either lenses or prisms. The type of refractor most commonly employed in indoor lighting is the lens found in tubular fluorescent-lamp luminaires intended for general lighting. It consists of a horizontal plastic panel which is mounted just below the lamp. The panel is flat on the top and has a special pyramidal (prism) or lens structure on the underside, which directs the light in certain directions and reduces the brightness under specific angles. Where in the past prismatic controllers were used, we see today more advanced lens-type refractors that give more



Luminaires, Fig. 5 Bending of light by refraction from the incident angle to the refraction angle

accurate possibilities to shape the light distribution. These types of refractors are also used to produce LED luminaires for general indoor lighting and for road lighting (Fig. 6). Refracting glass bowls were in the past sometimes used for high-pressure mercury and sodium road-lighting luminaires. They have become obsolete because they are heavy, but more so because lighting control in the upward direction, and therefore control of light pollution, is not easily attainable.

Diffusers

Translucent diffusers enlarge the apparent size of the light source. They scatter the light of the lamp in all directions without defining its light distribution. They serve mainly to reduce the brightness of the luminaire and thus the glare created by it. Diffusers are made of opal glass or translucent plastic, commonly acrylic or polycarbonate. The material should be such that it scatters the light while producing the minimum amount of absorption.

Screenings

Screening the Lamp from Direct View

Screening is employed to hide the bright lamp or lamps from direct view. The higher the brightness (luminance) of the lamp, the stricter the requirements for the shielding.



Luminaires, Fig. 6 Micro-lens type of refractor for a LED luminaire for general indoor lighting

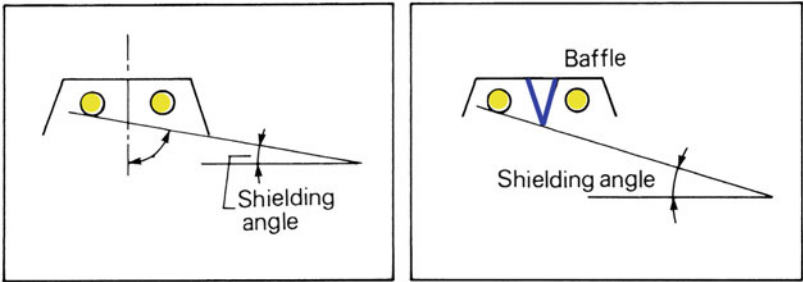
The luminaire reflector housing itself, or a built-in baffle, can provide the screening function (Fig. 7). When the sole purpose of the louvre is to shield the lamp from view, diffuse-reflecting material is used, such as a white-plastic louvre or, in the case of floodlights, matt-black metal rings (Fig. 8).

Shielding devices are often combined with the function of defining the light distribution, in which case highly reflective material is used for the louvre.

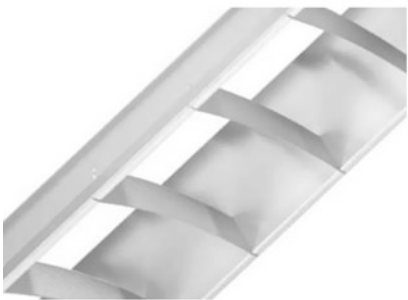
Color Filters

In certain lighting applications, in particular, display lighting and decorative floodlighting, color is sometimes used to help achieve the desired aesthetic effect. In the past, color filters attached to luminaires containing white light sources were extensively used for this purpose. Both absorption and dichroic (interference) filters were used, although absorption filters in particular (Fig. 9) lower the efficacy of the total lighting system. Typical transmittance values are as follows: for blue absorption filters 5%, for green absorption filters 15%, and for red absorption filters 20%. The consequence of the light absorption is that these filters become warm, which with high-power floodlights may damage the filter. The

Luminaires, Fig. 7 Lamp shielding by the reflector itself (*left*) and by an internal baffle (*right*)



Luminaires, Fig. 8 Simple louvre (*left*) to shield the lamp in a fluorescent-lamp luminaire and (*right*) a floodlight louvre



Luminaires, Fig. 9 Absorption-type color filters

solution where such floodlights are employed is to use dichroic filters, which are a more expensive alternative. Today, colored LED light sources are typically employed where colored lighting is

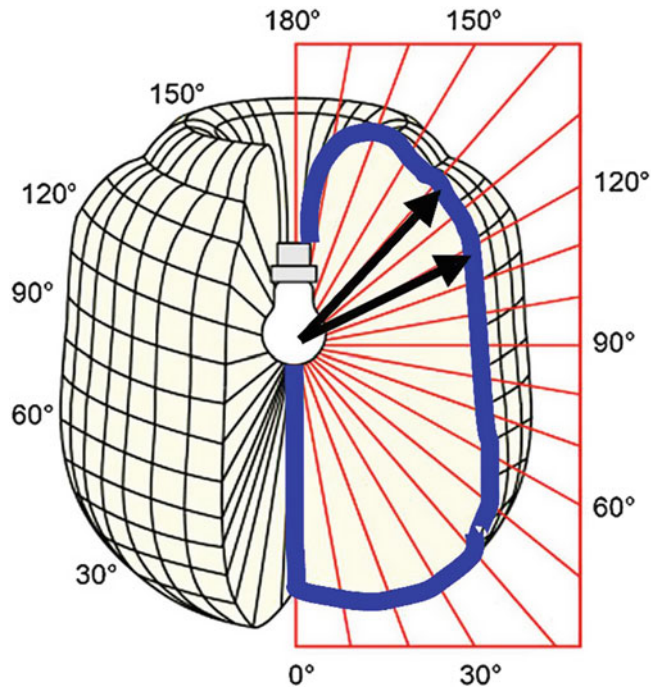
required. Here the color comes directly from the lamp itself, so the efficacy of the lighting system is much higher.

Light-Distribution Characteristics

The light distribution of a luminaire defines how the luminous flux radiated by the luminaire is distributed in the various directions within the space around it. It is the result of the optical devices used in the luminaire as described above. This is also called luminous intensity distribution, since it is specified in terms of luminous intensities in all the directions in which the luminaire radiates its light (Fig. 10). The luminous intensity diagram is, in fact, the light fingerprint of a luminaire, in digital form (I-table), and is the basis of all lighting calculations.

Basic photometric data that can be calculated from the light distribution are the beam spread and the luminaire light-output ratio. For all types of luminaires and for all types of application, these data provide an insight into the photometric quality of the luminaire.

Luminaires, Fig. 10 The light distribution of a luminaire given by its luminous intensity diagram. The arrows represent the luminous intensities in the directions specified. Here the light distribution is given for all planes, although it is usually only given for one (e.g., the *blue curve*) or two, mutually perpendicular, planes



Mechanical Characteristics

The mechanical function of the luminaire housing is threefold: it accommodates the various component parts of the luminaire, such as the optical system and the various components of the electrical system; protects these against external influences; and provides the means of mounting the luminaire in the installation.

Material

Sheet Steel

Sheet steel is generally chosen for the manufacture of tubular fluorescent luminaire housings for use indoors. The pre-painted sheet steel from the roll is white with diffuse reflection properties. Thus, after having been shaped in the luminaire factory into the desired luminaire form, no finishing-off operations are required.

Stainless Steel

Stainless steel is widely used for many of the small luminaire components, such as clips, hinges, mounting brackets, nuts, and bolts, that have to remain corrosion-free.

Aluminum Alloys

Aluminum alloys, in which other elements have been added to the pure aluminum to improve their mechanical, physical, and chemical protective properties, are used to manufacture cast, extruded, and sheet-metal luminaires. Cast aluminum refers to the process in which molten aluminum alloy is poured (cast) in a mold. Extrusion is the process in which softened aluminum alloy is pressed through the openings of a die. Cast and extruded aluminum alloys are much used in housings for floodlighting and road- and tunnel-lighting luminaires because they can be employed in humid and damp atmospheres without having to add protective finishes. Sheet aluminum is chiefly employed in luminaires for reflectors (Fig. 11). The reflectors are anodized to improve their reflection properties and to protect them from becoming matt.

Plastics

Plastics are used for complete luminaire housings, for transparent or translucent luminaire covers, and for many smaller component parts. All-plastic houses can of course only be employed for light



Luminaires, Fig. 11 Mirror reflector of sheet aluminum protected by a plastic film, which must be removed before use

sources that have a relatively low operating temperature.

Plastic covers are of methacrylate or polycarbonate. Methacrylate maintains its high light transmission properties over a long period, but its impact resistance is relatively low. The impact resistance of polycarbonates is very high and thus offers a high degree of protection against vandalism. It can be chemically treated to protect it from yellowing under the influence of ultraviolet radiation.

Glass

Although glass is heavy, glass covers are used where these have to be positioned close to a light source having a high operating temperature. This is the case, for example, with HID flat-cover road-lighting luminaires and with most floodlighting luminaires. Two sorts of glass are used:

- Regular glass, where no special demands are placed on heat resistance.
- Hard glass, where heat resistance, chemical stability, and resistance to shock are important. Should hard glass break, it will disintegrate into small pieces.

Luminaires made entirely out of glass are extremely heavy and nowadays are seldom employed.

Ceramics

Ceramic material is used in compact housings that are exposed to very high temperatures such as very compact halogen lamp luminaires.

Strength

All luminaires should have housings of sufficient rigidity to withstand normal handling, installation, and use. With indoor-lighting luminaires for fluorescent lamps, stiffness and rigidity of construction are particularly important, since these lamps are relatively large and awkward to handle. Perhaps the most critical part of a luminaire as far as strength is concerned are the mounting brackets. The strength required here is covered by a safety factor: the mounting bracket(s) must be able to support at least five times the weight of the luminaire itself. With road-lighting and outdoor floodlighting luminaires, the mounting brackets must also be strong enough to withstand the highest conceivable wind loading for the location. Here a good aerodynamic shape for the luminaire (characterized by its so-called shape factor) can be advantageous, as it also serves to reduce the strength required for the lighting mast.

Under some circumstances, the impact resistance of the luminaire itself is also important, particularly where protection against vandalism is called for.

Resistance to Pollution and Humidity

The atmosphere can contain many potentially corrosive gases which, in the presence of moisture vapor, will form highly corrosive compounds. In all areas where this danger exists (notably in outdoor applications, indoor swimming pools, and certain industrial premises), luminaires made from corrosion-resistant materials or having protective finishes should be used. In such areas, the luminaire should protect the optical and electrical

components it houses. It should, of course, be fully enclosed. The degree of protection provided by the luminaire is classified according to the International Protection Code (IP Code) as described in an IEC standard [2]. The IP Code consists of two numerals: IP • •.

- The first numeral classifies the degree of protection against the ingress of solid foreign bodies (ranging from fingers and tools to fine dust) and protection against access to hazardous parts.
- The second numeral classifies the degree of protection against the ingress of moisture.

The higher the IP values, the better the protection and thus, the lower the dirt accumulation and the lumen depreciation.

Ease of Installation and Maintenance

Many luminaires are of such a shape, size, and weight as to make mounting them a difficult and time-consuming operation. Mounting but also relamping and cleaning must usually be carried out high above ground level. So the ergonomic design of the luminaire should be such as to make these operations as easy and as safe as possible to perform. For example, covers should be hinged so

that the electrician has his hands free to work on the lamp and gear. A good, ergonomically designed luminaire is one that can be mounted in stages: first the empty housing or a simple mounting plate, which is light and easily handled, and then the remaining parts.

Electrical Characteristics

The electrical function of a luminaire is to provide the correct voltage and current for the proper functioning of the lamp in such a way as to ensure the electrical safety of the luminaire.

Lamp Holders

The most usual types of holders are the Edison screw, the bayonet, and the pin (Fig. 12). Most Edison screw and bayonet holders are made of plastics or porcelain, with metal parts for carrying the current. Porcelain is resistant to high temperatures and has a high voltage-breakdown resistance, which is important considering the high ignition voltage of HID lamps. The pin lamp holders for tubular fluorescent and compact fluorescent lamps are nearly always made of plastic. The metal contacts are spring-loaded to ensure constant contact pressure.

Luminaires, Fig. 12 A variety of lamp holders made of both plastics and porcelain



Electrical Wiring

The electrical wiring in a luminaire must be such as to ensure electrical safety. This necessitates great care in the choice of wire used and its installation. There are a great many different types of wire available, in both single-core (solid) and multicore (stranded) versions, all with various cross-sectional areas and clad with various thicknesses and qualities of insulation. The cross-sectional area (thickness) of the wire must be matched to the strength of the current flowing through it. The insulation of the wire used must be resistant to the high air temperature in the luminaire and the temperatures of the luminaire materials with which it is in direct contact. This is true not only under normal conditions of operation but also in the presence of a fault condition.

Mains Connection

The method used to connect a luminaire to the power supply must be both quick and safe. The practice generally adopted is to incorporate a connection block in the housing, although prewired luminaires in which the electrical connection to the mains is automatically made when the unit is placed in position are also available.

Electrical Safety Classification

The electrical safety classification drawn up by the IEC embraces four luminaire classes [2]:

- Class 0: Applicable to ordinary luminaires only. These are luminaires having functional insulation, but not double or reinforced insulation throughout and with no provision for earthing.
- Class I: Luminaires in this class, besides being electrically insulated, are also provided with an earthing point connecting all those exposed metal parts that could become live in the presence of a fault condition.
- Class II: This class embraces luminaires that are so designed and constructed that exposed metal parts cannot become live. This can be achieved by means of either reinforced or double insulation. They have no provision for earthing.
- Class III: Luminaires in this class are those designed for connection to extra-low-voltage circuits (referred to as safety extra-low voltage, SELV). They have no provision for earthing.

Thermal Characteristics

Temperature Control

A considerable amount of the electrical energy supplied to the lamp is converted into heat. The ballast adds to this heating effect within the luminaire. With very high-powered lamps, the ballast should be placed outside the luminaire in a special ballast box.

For a given lamp/ballast combination, the working temperature reached by the luminaire is dependent upon three factors:

- The volume of the luminaire. The greater the volume, the lower will be the temperature rise inside the luminaire.
- The ease with which the heat generated within the luminaire can be conducted through it to the surrounding air. One way of promoting airflow through the housing is to make use of heat-conducting materials in its construction. Most metals are suitable in this respect, while plastics, on the other hand, are thermal insulators and cannot, therefore, be employed as housing materials where high-power lamps are involved.
- The cooling effect of the surrounding air. Good heat dissipation calls for large surface areas to be in contact with the surrounding air. Luminaires for high-power lamps, such as high-bay luminaires, floodlights, and some LED luminaires that are very sensitive to high temperatures, are therefore provided with cooling fins (Fig. 13). Some types of industrial luminaires are provided with air vents in the top of the housing to allow the warm air to escape.



Luminaires, Fig. 13 Metal halide high-bay luminaire (*left*) and recessed LED downlight with cooling fins

Protection Against Flammability

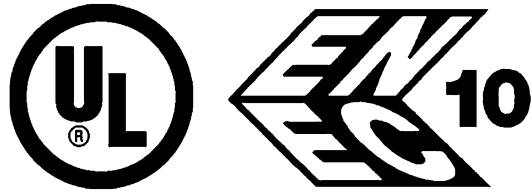
The flammability of a luminaire operating under fault conditions is an issue with luminaires made of plastics. But the combustion behavior of such luminaires is not just material dependent; it also depends on the shape and thickness of the luminaire housing.

Aesthetics

No less important than the functional characteristics of a luminaire is what is termed its aesthetical appeal, that is to say, its appearance and styling. In interiors, all non-recessed luminaires are clearly visible, and so whether switched on or not, their design should be in harmony with that of the interior. In outdoor lighting, it is usually only the daytime appearance of the luminaires, when these are clearly visible, that is important: particularly in built-up areas, their design can make a positive contribution to the attractiveness of the locality.

Approval

Luminaires always have to comply with the appropriate safety rules. UL (Underwriters Laboratories) is the US mark for demonstrating compliance with all US safety standards, and ENEC (European Norms Electrical Certification) is the similar European mark (Fig. 14).



Luminaires, Fig. 14 UL and ENEC certification marks. The number in the ENEC mark indicates the country of the institute that has given the European approval

Cross-References

- ▶ [Interior Lighting.](#)
- ▶ [Light Distribution.](#)
- ▶ [Light Pollution.](#)
- ▶ [Lumen Depreciation.](#)
- ▶ [Road Lighting.](#)
- ▶ [Sports Lighting.](#)

References

1. CIE Publication: S 017/E: 2011, International lighting vocabulary. International Lighting Commission, CIE, Vienna (2011)
2. IEC International Standard, 60598-1: ed 7.0, Luminaires – part 1: General requirements and tests. (2008)

Luminance Measurement Device

- ▶ [Luminance Meter](#)

Luminance Meter

Bor-Jiunn Wen¹ and Chao-Hua Wen²

¹Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Luminance measurement device](#)

Definition

A luminance meter is a device used to measure the photometric unit, luminance, in a particular direction at a solid angle from a surface. There are three methods to measure luminance. The first method is the conventional luminance meter that uses a designed sensor with a spectral response that closely matches the $V(\lambda)$ spectral luminous efficiency function of the human eye. The second method is the measured luminance that is derived through the illuminance measurement and subsequent calculations. The third method is the imaging luminance measurement device that measures the luminance in a single direction and is similar to a digital camera which records color images.

Introduction

The luminance of a light source (integrating sphere with known output aperture) is determined from the measurement geometry and illuminance measured by a photometer. The luminance of the light source is obtained as Eq. (1),

$$L_v = \frac{E_v D^2}{A} \quad (1)$$

where E_v is the illuminance at a distance D between the aperture plane of the source and the photometer and A is the area of the source aperture. The distance D depends on the radius of the limiting aperture r_1 , the radius of the source r_2 , and the physical distance d between the source and the aperture according to Eq. (2):

$$D^2 = r_1^2 + r_2^2 + d^2 \quad (2)$$

Equation (2) is accurate within $\pm 0.01\%$ for distances that are more than one decade greater than the radius [1].

Since a regular digital camera does not equip the photopic filter, its pixels are specified in a device-independent RGB color space in general. The photometric data can be calibrated from a color image through a series of calculations that involve conversion from RGB to CIE's absolute

color space pixel-by-pixel. In addition, the human vision system can adapt to luminance in a very wide range (10^{-6} – 10^6 cd/m²), but the digital image's dynamic range is 8 bit or less than 12 bit. Therefore, in order to measure a high dynamic scene, it is necessary to know the exposure values of the digital camera as Eq. (3):

$$EV = \log_2 \left(\frac{f^2}{t} \cdot \frac{S}{100} \right) \quad (3)$$

where f is aperture, t is shutter speed, and S is ISO sensitivity value. The dynamic range is just adjusted by shutter speed if both the aperture and the ISO value are fixed during the measurement. The processes of calibration for the imaging luminance measurement device are the procedure for determining the mathematical transform between luminance and the digital data L of image with shutter speed t [2].

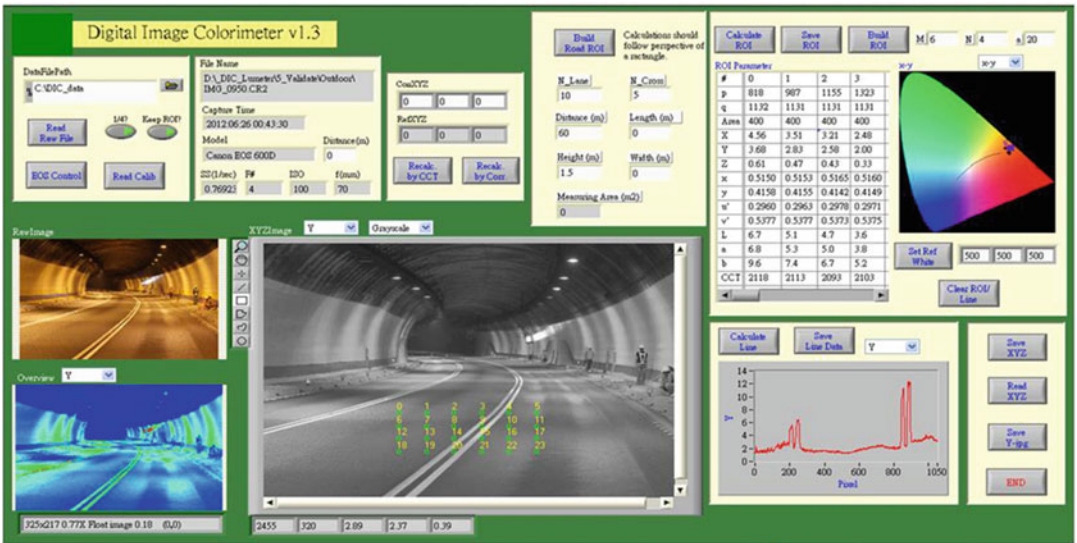
Applications

An imaging luminance measurement device (ILMD) is capable of measuring illuminance in an image. It can automatically quantify the distribution or the uniformity of light. This is much more powerful than the conventional single-point luminance meter. For example, an ILMD is utilized for measuring the luminance of all pixels of flat panel displays [2]. Figure 1 depicts the measurement of the luminance of vertical or horizontal test patterns shown on a TFT-LCD display with an ILMD in a slightly tilted position [2]. The subsequent image processing includes calculating the spatial frequency response (SFR) of the image using a slanted-edge algorithm, obtaining resolution at a specified decrease (e.g., 50%) in the SFR of the device under test (DUT), and finally, specifying the resolution of the DUT in terms of the image sharpness.

In addition, an automatic measurement system based on an ILMD also measures the photometric properties of road and tunnel lighting systems [3, 4]. Figure 2 depicts an automatic system for this purpose. Compared with a single-point luminance

Luminance Meter,

Fig. 1 Sharpness measurement of FPD by using a ILMD



Luminance Meter, Fig. 2 Automatic measurement and analysis program for road and tunnel lighting systems

meter (conventional luminance meter), the ILMD is more precise and time-saving for measuring road and tunnel lighting.

Since the rapid growth of light-emitting diode (LED) technologies, the adjustable lighting on billboards using LED has become increasingly popular. Many of these products are pushed to much higher contrast in spatial and/or temporal configurations to attract attention. However, the glare and/or flicker of the LED sources may produce an uncomfortable visual experience.

Therefore, Hsu et al. [3] proposed a measurement system for regulating between flicker and glare by using perceptual ratings of LED billboards under various conditions. Authors performed both objective and subjective evaluations of flashing LED billboards in interior spaces. The objective flicker and glare evaluations were carried out by taking temporal and spatial measurements, respectively. The properties of the flicker and glare values obtained are reproduced under physical conditions. The visual results were modelled

by simple equations as a function of objective measurements in terms of a low-pass flicker index and unified glare rating. Thus, an effective method was developed for regulating the degree of flashing LED lighting using an ILMD.

Cross-References

- [Illuminance Meter](#)
- [Standard Measurement Geometries](#)

References

1. Kostkowski, H.J.: Reliable Spectroradiometry. Spectroradiometry Consulting, La Plata (1997)
2. SID/ICDM: Information Display Measurements Standard, v1.03 (2012)
3. Hsu, S.W., Chung, T.Y., Pong, B.J., Chen, Y.C., Hsieh, P.H., Lin, M.W.: Relations between flicker, glare, and perceptual ratings of LED billboards under various conditions. In: CIE Centenary Conference, Paris (2013)
4. CIE 194: On Site Measurement of the Photometric Properties of Road and Tunnel Lighting (2011)

Luminescence

Joanne Zwinkels
National Research Council Canada, Ottawa, ON,
Canada

Synonyms

[Cold body radiation](#); [Cold light](#)

Definition

Luminescence is the emission of optical radiation caused by a non-thermal process [1]. Luminescence is therefore a form of “cold body radiation” or “cold light,” since it does not result from heating (i.e., from a relatively cool material, not above room temperature). It is therefore distinguished from light emitted by *incandescence*,

which is a form of blackbody radiation from a heated material, that is, “hot light” whose spectral power distribution is determined by its thermal temperature.

The term *luminescence* is also often used to generally discuss *fluorescence* and *phosphorescence* phenomena. However, this is only one type of luminescence, known as *photoluminescence*, caused by the absorption of optical radiation, and there are many other types of luminescence (e.g., cathodoluminescence, electroluminescence, chemiluminescence, bioluminescence, triboluminescence, thermoluminescence, etc.) classified by their different modes of activation, which are described below.

Overview

There are many early reported observations of the phenomenon of “luminescence.” For example, it was reported in the early seventeenth century by an Italian cobbler, Vincenzo Cascariolo, when he was searching for the philosopher’s stone by heating a mixture of coal and stones containing barium sulfate and found that, after cooling, it emitted a bluish glow in the dark for several hours [2]. The term “*phosphorescence*,” named for the luminous glow over the element phosphorus, is now frequently used for this phenomenon of long-lived or delayed luminescence. However, the term “luminescence” was first introduced as “*luminescenz*” in 1888 by the German physicist, Eilhard Wiedemann when he coined this term to describe the general phenomenon of light that did not depend upon the temperature of the substance concerned [3]. The term comes from the Latin root “*lumen*” meaning “light” plus “-escence.”

The process of luminescence occurs when an atom, molecule, crystal, or polymer is non-thermally excited to an excited electronic or vibrational state. When this excited specimen drops back down to its equilibrium state, it emits radiation if it is not in thermal equilibrium with its environment. The emitted radiation is usually in the visible spectral region but it can also be in the ultraviolet or infrared spectral regions and is generally of lower energy, that is, *downconverted* (longer wavelength) compared to the excited

radiation. This excitation of luminescence can be caused by a wide variety of mechanisms, with many technologically important applications, which are described below.

Types of Luminescence and Their Applications

The main categories of specialized luminescence and their mode of activation energy are [2, 4]:

Electroluminescence (EL) – action of an electric field in a gas or in a solid material

Cathodoluminescence (CL) – caused by the impact of an incident electron beam on certain types of luminescent materials, such as the coating of a television screen

Photoluminescence (PL) – caused by absorption of optical radiation

Radioluminescence (RL) – caused by X-rays or radioactive radiation

Chemiluminescence (ChL) – caused by energy released by a chemical reaction, usually oxidation

Bioluminescence (BL) – caused by chemiluminescence occurring in living organisms

Triboluminescence or mechanoluminescence (ML) – caused by the action of mechanical forces

Thermoluminescence (TL) – occurs when a previously excited luminescent material is heated

This is not an exhaustive list but illustrates the wide variety of luminescent mechanisms. Of these different types of specialized luminescence, *electroluminescence* is perhaps the most technologically important since a light emitter has to be able to produce light in an efficient manner under electrical excitation [5]. The passage of electric current through a gas can produce light as in gas lasers, sparks, arcs, lighting, coronas, and electrical-discharge lamps such as deuterium and xenon. Light production by direct electroluminescence occurs when a beam of electrons is accelerated and impinges on a phosphor screen. This is known as *cathodoluminescence*, and the best-known example is the technology used in a cathode ray tube

(CRT) color television screen, where three electron beams striking an array of red, green, and blue-emitting crystal phosphors, can lead to a full range of colors. The cathodoluminescence technique induces the emission of light of characteristic wavelengths in a wide spectral range. This is utilized in the scanning electron microscope (SEM)-CL to obtain chemical composition information and spatial details of semiconducting materials and for detecting trace rare-earth elements [5].

Luminescence from phosphors, which are based on wide band-gap materials containing activators, which cover the entire visible range, is used in a number of materials and devices, including fluorescent lamps and tubes, displays, and light-emitting diodes (LEDs) [6]. LEDs, which are solid-state semiconductor devices, have a “tunable color” whose emission wavelength range depends upon the properties of the doping impurities that are used to form the electrical junction. For this reason, photoluminescence is commonly used for the analysis of the optical behavior of impurities in semiconducting materials for their use in optoelectronic devices [5]. In 1966, a phosphor coating was applied to a blue LED to create the world’s first white LED. In the past 20 years, this technology has shown a phenomenal rate of improvement in efficiency, rapidly replacing traditional incandescent and fluorescent lamps for general lighting applications.

For photoluminescence-based lighting to be efficient, the number of photons (light) that is emitted per photons absorbed by the material (*photoluminescence quantum yield*) needs to be high. Electroluminescent quantum dot (QD) displays, which are a self-emissive technology that does not depend on inefficient down-conversion of blue photons to red and green photons but electrically stimulates the QDs to emit photons, promise dramatic improvements in efficiency over photoluminescent QD displays because there is no wasted light [7]. As mentioned above, photoluminescence can be further subdivided into two forms: *fluorescence* and *phosphorescence*. *Fluorescence* is due to transitions between states of the same multiplicity (i.e., allowed transitions) and the emitted radiation generally appears within 10 nanoseconds of

excitation. *Phosphorescence*, on the other hand, is due to transitions between states of different multiplicity (i.e., forbidden transitions) and the emission of optical radiation appears 10 nanoseconds or longer after the excitation. However, for most types of optical radiation detectors used in photoluminescence spectroscopy, for example, photomultipliers and silicon photodiodes, the time constant of the detector is too long in order to be able to distinguish between these two forms of photoluminescence, and the term fluorescence is commonly used for the measurement of both phenomena.

Radioluminescence, also known as *scintillation*, is induced by the various “rays” emitted by ionizing radiation interacting with radioactive substances. This includes alpha rays (helium nuclei), beta rays (electrons), and gamma rays (high-energy electromagnetic radiation). A remarkable natural occurrence of radioluminescence is the aurora borealis, in which radioactive processes in the sun, excited by the solar wind, produce large numbers of electrons and ions that emit into space. Self-luminous paints that are used in dial markings on watches and other instruments are made of a radioactive substance, for example, tritium or radium, hitting a crystal phosphor, and producing tiny “*scintillations*” that are visible to the human eye.

Chemiluminescence is produced by chemical reactions, mainly oxidation, at low temperatures. Commercial forms of chemiluminescence include emergency glowsticks and glowing liquid-filled plastic tubes.

Bioluminescence is light produced from a chemical reaction involving a chemical substance in a living organism and is a form of the more general chemiluminescence. It occurs in glow-worms, fireflies, fungi, putrefied meat, and in many oceans or sea creatures, such as jellyfish and deep-sea water fishes.

Triboluminescence is the production of light by mechanical disturbances on certain crystalline substances. This can include friction and static forces such as crushing, grinding, cutting, etc. For example, the crushing of sugar crystals or the cutting of ruby or sapphire crystals with a diamond saw can produce light.

Thermoluminescence and optically stimulated luminescence (OSL) are used in luminescence-based dating, for example, on the planet Mars for dating sediments and in fluvial deposits in order to determine the age of sediments deposited during the last glacial-interglacial cycle. These techniques are also used in nuclear-accident dosimetry and in the investigation of luminescent processes [8].

Luminescent Materials

Luminescent materials can be classified into several broad groups. These are summarized in Table 1, along with some specific examples and typical applications [4].

Fluorescent whitening agents (FWAs) or optical brightening agents (OBAs) are an important class of organic luminescent materials. These are

Luminescence, Table 1 Examples of luminescent materials and their applications

Classification	Examples	Applications
Aromatic molecules	Xanthenes, aminocoumarins, stilbenes, and oxazines	Scintillators, luminescent dyes/paints, FWAs, and dye lasers
Inorganic crystals	Diamond, ruby, and alkali halides	Scintillators, solid-state lasers, and jewelry
Noble gases	He, Ne, Ar, Kr, and Xe	Discharge lamps, gas lasers, and scintillators
Inorganic ions	Rare earths	Nd glass lasers and glass scintillators
Simple inorganic molecules	N ₂ , I ₂ , and CO ₂ H ₂ , D ₂ , N ₂ , and Hg	Gas lasers Discharge lamps
Biological molecules	Amino acids, nucleotides, Vitamins, and hormones	Biological research
Aliphatic molecules	Paraffins and cyclohexane	Emit in far-UV

generally sulphonated stilbenes that absorb short-wave radiation in the ultraviolet and violet range (below 430 nm) and emit blue light with a peak wavelength around 450–470 nm, producing a bluish tint, which increases our perception of whiteness significantly, so that the FWA-containing material appears to be “whiter than white.” These FWAs are used in the manufacture of white products such as paper and textiles and are used in laundry detergents, some plastics, and cosmetics.

Photoluminescent inorganic pigments are used in diverse ways, including outdoor displays and in the manufacture of safety goods including outdoor safety signing and marking and safety clothing. These pigments absorb daylight radiation in the UV or short-wave visible and emit at longer wavelengths in the green-yellow or yellow-orange visible range, which enhances their perceived colorfulness and visibility.

Luminescence in inorganic crystals, such as ruby and diamond, is due to the presence of impurity centers or crystal defects. This process can be very efficient and is used in the manufacture of solid-state lasers. The noble gases, such as He, Ne, Ar, Kr, and Xe, are often used in low-pressure discharge lamps, which emit light of different colors and are used in advertising installations. On the other hand, low-pressure discharge lamps made with mercury vapor and phosphor (e.g., fluorescent lamps), are suitable for general lighting purposes. High-pressure discharge lamps are usually made using Hg or other elements combined with Hg and can also produce significant amounts of thermal radiation, that is, not only luminescence. One of their main applications is for the illumination of large indoor or outdoor spaces, such as sports stadiums, and parking lots.

Luminescent solid-state materials have largely displaced gas discharge luminescence in many applications because of the high emission efficiencies demonstrated by both organic light-emitting diodes (OLEDs) and inorganic light-emitting diodes. Other solid-state luminescent materials and devices of current interest include down-conversion materials, nanomaterials and powder, and thin-film electroluminescent phosphor materials and devices. Lanthanide-doped

nanoparticles are an emerging class of luminescent nanomaterials that are of considerable recent interest for both biological sciences (e.g., use as biological labels) and photovoltaic (PV) applications because of their low toxicity, high chemical stability, and upconversion properties, that is, potential to convert light between different spectral regions with high signal-to-noise (S/N) sensitivity [7].

Photoluminescence is the most widely used technique for the analysis of semiconducting materials because it is much more sensitive than reflectance or absorbance measurements for detecting impurities at low concentrations [5]. Because of its high sensitivity, intrinsic selectivity, non-invasive nature, and potential for multiplexing applications, luminescence techniques are one of the most widely used detection methods in biomedical science. Fluorescence spectroscopy is also an important tool in biological research involving structure elucidation and the study of interactions in proteins and nucleic acids [9].

Cross-References

- [Blackbody and Blackbody Radiation](#)
- [Incandescence](#)
- [Light, Electromagnetic Spectrum](#)

References

1. CIE S 017/E: *ILV – International lighting vocabulary*, 2nd edn. CIE Central Bureau, Vienna (2020)
2. Nassau, K.: *The physics and chemistry of color: the fifteen causes of color*. John Wiley & Sons, New York (1983)
3. Birks, J.B.: *Standardization in spectrophotometry and luminescence measurements*, NBS Special Publication 466 (edited by K.D. Mielenz *et al.*) Gaithersburg (1977)
4. Mielenz, K.D.: *Photoluminescence spectroscopy, optical radiation measurements*, Vol. 3. Academic, Orlando (1982)
5. Bass, M.: *Handbook of optics, devices, measurements & properties*, vol. II, 2nd edn. McGraw-Hill Inc., New York (1995)
6. Kuehni, R.G.: *Color – An introduction to practice and principles*, 3rd edn. John Wiley & Sons, New York (2012)
7. Kitei, A.H.: *Solid state luminescence – Theory, materials and devices*. Chapman and Hall, London (1993)

8. Botter-Jensen, L., Thomsen, K.J., Jain, M.: Review of optically stimulated luminescence (OSL) instrumental developments for retrospective dosimetry. *Radiation Measurements*. **45**, 253–257 (2010)
9. Kitei, A.: *Luminescent materials and applications*. John Wiley & Sons Ltd, West Sussex (2008)

Luminous Intensity Distribution

- [Light Distribution](#)

Luminescent Materials

- [Phosphors and Fluorescent Powders](#)

Luminous Intensity Meter

- [Instrument: Photometer](#)

Luminous Efficiency Function

- [Spectral Luminous Efficiency](#)

Luxmeter

- [Illuminance Meter](#)

M

M, P, and K Pathways

► [Magno-, Parvo-, and Koniocellular Pathways](#)

MacAdam, David L

Michael H. Brill¹ and Rolf G. Kuehni²

¹Datacolor, Lawrenceville, NJ, USA

²Charlotte, NC, USA



David Lewis MacAdam (1910–1998) was an American physicist and color scientist who made important contributions to colorimetry, color discrimination, color photography and television, and color order.

MacAdam grew up in Philadelphia, attended Lehigh University, and in 1936 received a PhD in physics from MIT. Under Prof. Arthur C. Hardy, he originated the first course in color measurement and assisted Hardy in the preparation of *Handbook of Colorimetry*, published in 1936.

Upon graduation, MacAdam joined the Research Laboratories of the Eastman Kodak company in Rochester, NY, from where he retired as a Senior Research Associate in 1975. Subsequently, he was named Adjunct Professor at the University of Rochester, Institute of Optics, where he remained active until 1995. At Eastman Kodak, among many other things, he helped to establish the theoretical basis for color photography, including color masking as compensation for unwanted dye layer absorptions [1].

Optimal Object Color Limits

While still studying, MacAdam published in 1935 two papers in which he offered a geometric proof of the optimal-color theorem and calculated the optimal object color solid raised over the CIE chromaticity diagram, using the newly established CIE standard observer and illuminant C and A data from 1932 [2, 3].

MacAdam Ellipses

One of MacAdam's best-known contributions was in support of technological color control. Assuming that the basis of color difference perception was the statistical error in matching the appearance of a given color stimulus, he conducted an extensive experiment with one observer, the result of which was expressed in the CIE chromaticity diagram in form of statistically derived ellipses, published in 1942 [4]. However, the resulting Friele–MacAdam–Chickering color difference formula proved less effective in predicting perceived color differences than formulas derived on other bases. In attempting to convert the ellipses to circles of equal size, MacAdam encountered the non-Euclidean nature of psychophysical color space.

Instrumentation and Computation

In the mid-1940s MacAdam pioneered the use of computers in colorimetric computations, established Hardy's reflectance spectrophotometer as a reliable industrial measuring instrument, and invented a tristimulus integrator as an accessory.

Principal-Component Analysis of Daylight

With Deane B. Judd and Günter Wyszecki, MacAdam performed the first principal-component analysis of phases of daylight of various correlated color temperatures, demonstrating that they can be represented as linear combinations of a limited number of spectral components [5].

Optical Society of America Uniform Color Scales

MacAdam was a leading member of the committee of the Optical Society of America that in 1947, at the suggestion of the US National Research

Council, began work on a perceptually uniform colorimetric model of the color solid. Upon the retirement of its first chairman D. B. Judd, MacAdam was elected chair. The result of the committee's work was published in 1974 as "Uniform Color Scales," [6] with OSA publishing a related color atlas with 558 samples in 1977.

MacAdam was a major contributor to OSA's 1953 book *The Science of Color* [7]. His interest in the history resulted in publication in 1970 of *Sources of Color Science* [8], a compilation of 26 seminal papers on color science, from Plato to Le Gros Clark, several translated for the first time into English. In 1981 he published *Color Measurement: Theme and Variations*, a presentation of fields of color science in which he made important contributions [9]. He also authored some 100 peer-reviewed journal articles.

MacAdam was president of the Optical Society of America in 1963 and editor of the *Journal of the Optical Society of America* from 1964 to 1975. He was active in the Inter-Society Color Council and the Commission Internationale de l'Eclairage (CIE). He received honors from many societies, including the Frederic Ives Medal of the Optical Society of America in 1974.

References

1. MacAdam, D.L.: Subtractive color mixture and color reproduction. *J. Opt. Soc. Am.* **28**, 466–480 (1938)
2. MacAdam, D.L.: The theory of maximum visual efficiency of colored materials. *J. Opt. Soc. Am.* **25**, 249–252 (1935)
3. MacAdam, D.L.: Maximum visual efficiency of colored materials. *J. Opt. Soc. Am.* **25**, 361–367 (1935)
4. MacAdam, D.L.: Visual sensitivities to color differences in daylight. *J. Opt. Soc. Am.* **32**, 247–274 (1942)
5. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1036 (1964)
6. MacAdam, D.L.: Uniform color scales. *J. Opt. Soc. Am.* **64**, 1591–1702 (1974)
7. Optical Society of America Committee on Colorimetry: *The Science of Color*. Crowell, New York (1953)
8. MacAdam, D.L.: *Sources of Color Science*. MIT Press, Cambridge, MA (1970)
9. MacAdam, D.L.: *Color Measurement: Theme and Variations*. Springer, New York (1981)

Mach Bands

Rolf G. Kuehni
Color Science Laboratory, Carolina State
University, Raleigh, NC, USA

Definition

Mach bands are brightness illusions, originally referring to a situation where a lighter and a darker sharp-edged gray area are separated from each other by a gray area where the brightness transforms uniformly from the brighter to the darker field. When looking at an image producing perceived Mach bands (Fig. 1), two bands are seen, a lighter one between the light and the intermediate fields and a darker one between the intermediate and the dark fields, even though the physical composition of the three fields does not contain physical sources of the bands. More recently, the term Mach bands is also applied in the case of sharply adjacent gray fields of different reflectance, each band physically uniform, where, in addition to the perceived change in lightness across each band, very narrow lighter and darker bands may be perceived near the line where two fields connect (Fig. 2).



Mach Bands, Fig. 1 Image of classical Mach effect. Section **a** is a light gray field with uniform reflectance; from **b** to **c** there is a uniform reflectance transition from the color of section **a** to section **d** which is a reflectively uniform dark gray field; the light and dark perceptual Mach bands are located at **b** and **c**, respectively

History of Mach Bands

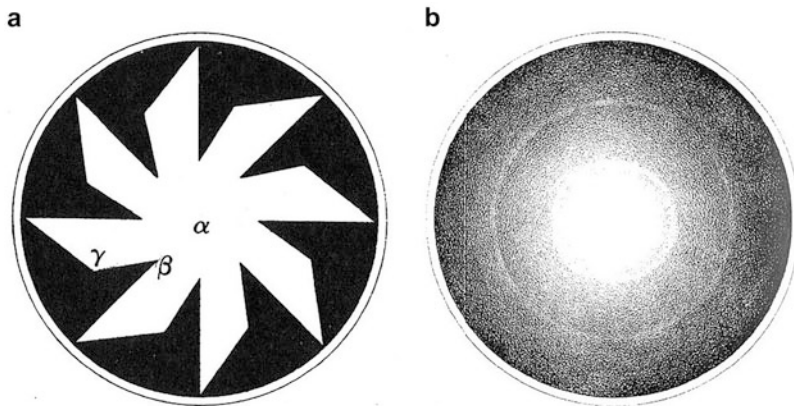
Perceptual band effects were discovered by the Austrian chemist Ernst Mach (1838–1916) in 1865 based on the appearance generated when disks that have different white and black patterns on them are rapidly rotated, an example of which is shown in Fig. 3a. Figure 3b is Mach's representation of the appearance when looking at the rapidly spun disk of Fig. 3a. The inner bright circle is perceptually experienced at the location β in Fig. 3a, and the outer bright circle at location γ . When black and white in Fig. 3a are reversed, the perceptual result of rapid rotation is dark perceptual bands at locations β and γ . Mach reported his findings of the perceptual illusion bands in five articles published between 1865 and 1868, only the first two of which are referenced here ([1, 2] English translations in F. Ratliff, 1965, [3]).

In later experiments, Mach thought of disk designs that produced both light and dark bands from the same spinning disk. He also produced Mach bands in square format images by fastening the related multiple identical images onto a cylinder that he then rapidly rotated.

The size of the distance between β and γ in Fig. 3b, comparable to the distance between **b** and **c** in Fig. 1, is an indicator of the perceptual intensity of the resulting Mach bands. The narrower the reflectance transition sector is (the closer to vertical the line between **b** and **c** in Fig. 1), the more perceptually strong the related Mach bands are. An example of the relationship between physical stimuli and perceptual results is shown in Fig. 4 [4]. The solid black line represents the physical measure of luminance (represented in foot-lambert), proceeding at two points from constant value to transitioning values. The black line with open circles represents average perceived lightness before and after the transition points, indicating the dark Mach band on the left and the light one on the right. With the rising solid line of luminance being relatively steep, the resulting light Mach band is perceived as particularly light. However, as the transition sector approaches or is zero in width the perceived Mach bands become very weak.



Mach Bands, Fig. 2 Five rectangular fields, each of which is physically uniform but changes in perceptual appearance from left to right. For some viewers, very narrow adjacent light and dark Mach bands are visible where fields touch



Mach Bands, Fig. 3 (a) (left) and (b) (right). Both images are from Mach's original article [3]. The left figure is an image of one of the circular disks that Mach rapidly rotated to demonstrate the phenomenon. Figure 3b is Mach's representation of the appearance of the rapidly

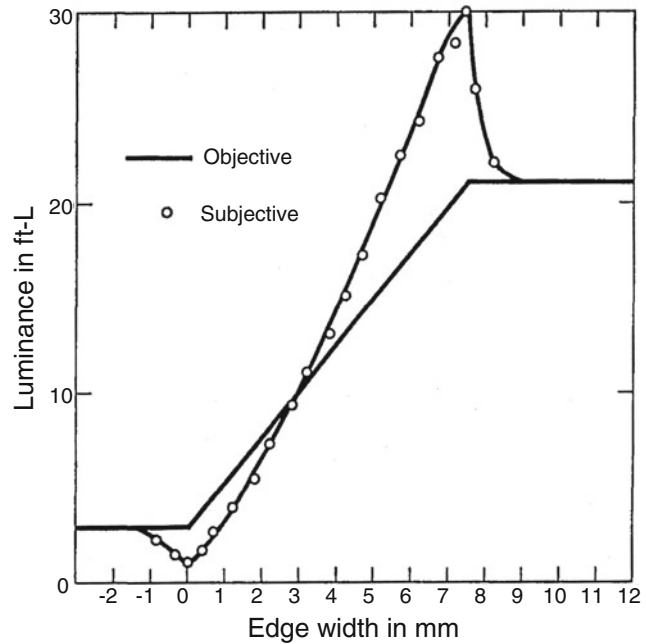
rotated disk shown in Fig. 3a. The inner perceived white Mach band appears at location β ; and the middle white band at location γ on the disk. Copyright permission unattainable (publisher no longer exists and author passed away)

The term Mach bands is now also applied when there is no transition field but the lighter and the darker fields are adjacent. In this situation, the perceived bands, a lighter one on the edge of the light field and a darker one on the edge of the dark field were reported in 1861 by the French Chemist M. E. Chevreul (1786–1889), known as the Chevreul illusion [5]. It is experienced when placing homogenous bands of different brightness/lightness adjacent to each other (Fig. 2). Even though the bands are physically uniform they are perceived as not so, with the appearance being lighter where the field is adjacent to a darker

one and vice versa. The effect of the Chevreul illusion has been enlarged in the Craik-O'Brien-Cornsweet illusion [6] by placing a narrow band of increased lightness at the edge of the light field and one of decreased lightness at the adjacent darker band. This physical change enlarges the perceptual effect of the Chevreul illusion. It does not play a role in the definition of Mach bands.

The classical Mach bands have been experimentally investigated since their discovery. In 1972, A. Fiorentini surveyed research pertaining to the Mach band phenomena since the 1950s [7]. This included determining the effect of the

Mach Bands, Fig. 4 A comparison between the physical dimensions of a Mach band stimulus (solid line) and the average perceived perceptual results from a number of observers (line with open circles), indicating the location of the dark (lower bulge) and the light (upper bulge) peaks of the resulting perceived Mach bands [4]. Reprinted with permission from the copyright holder The Optical Society of America (OSA)



steepness of the transition from low to high light levels (examples of such research are found in [8–10]). Substantial inter-observer variability in the perception of the Mach band effect has been discovered by multiple authors over the decades.

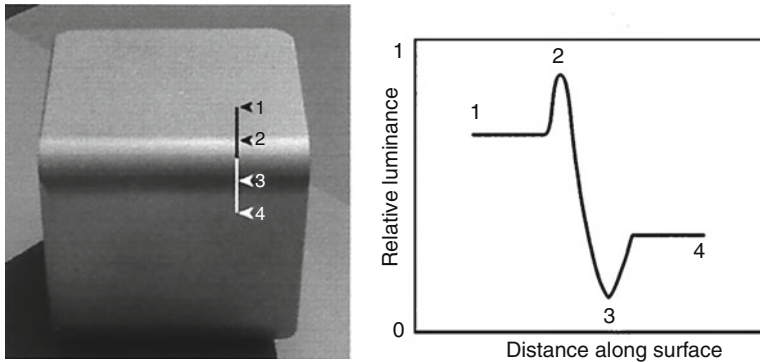
Mach bands are related to achromatic fields or to the reflectance differences of chromatic fields with hue and chroma having no apparent effect on their perception. An empirical interpretation of classical Mach bands was offered in 1999 by R.B. Lotto, S. M. Williams, and D. Purves [11]. They believe that Mach bands may be the result of the response of the human visual system to the relative frequency of certain reflectance patterns generated from light reflected from objects with graded surfaces. They offered as an example an image of a metal cube with some rounded edges, illuminated by sunlight (Fig. 5). The relative luminance of the reflected light from points 1 to 4 in the image, to a degree, duplicates that of the classical Mach bands. According to these authors, the human perceptual system is presumed to respond to a relatively high frequency of occurrence of such visual situations in nature. This hypothesis has not found general acceptance by sensory biologists.

An unresolved major issue regarding Mach bands is at what level in the vision process they

are being generated. It was initially presumed that Mach bands are generated in the retina but soon after, more complex processes in the visual system were also proposed. A detailed, scientifically solid, neural model does not yet exist. Early explanation attempts have been reviewed by F. Ratliff in his 1965 book *Mach Bands* (3). An interesting theory based on response normalization was proposed by F. A. A. Kingdom in 2014 which states, "... whereby the gains of early visual channels are adjusted on a local basis to make their responses more equal" [12].

Summary

Mach bands are perceptual illusions created in the brain/mind from neurobiological information generated in the retina when viewing stimuli of adjacent fields of different lightness, either separated by an intermediate field with reflectance changing from that of the lighter field to that of the darker field or by viewing two directly adjacent fields of different reflectance. While best visible in achromatic fields, chromaticness itself does not affect the Mach band effect. The complete neural basis of the perception of Mach bands has so far not been discovered.



Mach Bands, Fig. 5 left: An image of a metal cube with a rounded corner illuminated by sunlight with the locations of the light (1) and dark (4) surfaces and the locations of the light (2) and dark (3) Mach bands as interpreted by Lotto,

Williams, and Purves and shown in the conceptual image on the right (11). Reprinted with permission from the copyright holder PNAS (Proceedings of the National Academy of Sciences of the United States of America)

References

1. Mach, E., Ueber die Wirkung der räumlichen Verteilung des Lichtreizes auf die Netzhaut, Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften, Wien, **52/2**, pp. 303–322 (1865). English translation: F. Ratliff, Mach Bands, pp. 253–271 (1965) (Ref. 3)
2. Mach, E. Ueber den physiologischen Effect räumlich vertheilter Lichtreize, Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften, Wien, **54/2**, pp. 131–144 (1866). English translation: F. Ratliff, Mach Bands, pp. 272–284 (1965) (Ref. 3)
3. Ratliff, F.: Mach Bands: Quantitative Studies on Neural Networks in the Retina. Holden-Day, San Francisco CA (1965)
4. Lowry, E.M., DePalma, J.J.: Sine-wave response of the visual system I. the Mach phenomenon. J. Opt. Soc. Am. **51**, 740–746 (1961)
5. Chevreul, M.E.: Contrast of Colour and their Application to the Arts, English Translation by J. Spanton. Routledge, Warne and Routledge, London (1861) page 4, figs. 1 and 3
6. Cornsweet, T.N.: Visual perception. Academic Press, New York (1970)
7. Fiorentini, A.: Mach Band Phenomena, Chapter 8 in D. Jameson and L. M. Hurvich, Visual Psychophysics, vol VII/4 of Handbook of Sensory Physiology. Springer Verlag, Berlin, New York (1972)
8. Thomas, J.P.: Brightness variations in stimuli with ramp-like contours. J. Opt. Soc. Am. **56**, 238–242 (1966)
9. von Békésy, G.: Brightness distribution across the Mach bands measured with flicker photometry, and the linearity of sensory nervous interaction. J. Opt. Soc. Am. **58**, 1–8 (1968)
10. Buckle, C.E., Udawatta, V., Straus, C.M.: Now you see it, now you don't: visual illusions in radiology, Radiographics **33**, 2087–2102 (2013). The variation of the luminance gradient. Atti Fond. G. Ronchi. **13**, 145–155 (1958)
11. Lotto, R.B., Williams, S.M., Purves, D.: An empirical basis for Mach bands. Proc. Natl. Acad. Sci. U. S. A. **96**, 5239–5244 (1999)
12. Kingdom, F.A.A.: Mach bands explained by response normalization. Front. Hum. Neurosci. **9**, 943 (2014)

Madder Lake

► [Alizarin](#)

Magno-, Parvo-, and Koniocellular Pathways

Jasna Martinovic

School of Psychology, University of Aberdeen, Aberdeen, UK

Synonyms

[Geniculate pathways](#); [M, P, and K pathways](#); [Parallel visual pathways](#); [Retinogeniculate pathways](#); [Retinogeniculocortical pathways](#)

Definition

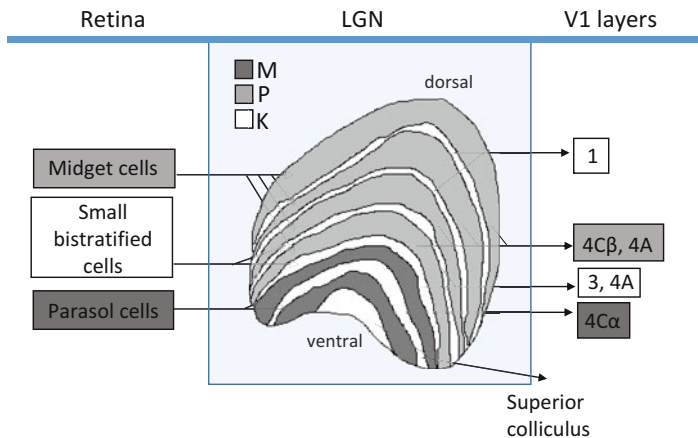
Magno-, parvo-, and koniocellular pathways are the three visual pathways in primates. These pathways are established at the level of the lateral geniculate nucleus (LGN) of the thalamus. They are formed of morphologically distinct cellular layers that receive information from different types of retinal ganglion cells and project to different layers in the primary visual cortex.

Anatomical Considerations

The LGN layers of each of the three visual pathways have a specific cytoarchitectonic structure. The names of the pathways are derived from these structural characteristics [1, 2]. Magnocellular (M) cells have relatively large bodies (lat. *Magnus*: large), parvocellular (P) cells have smaller bodies (lat. *Parvus*: small), while koniocellular (K) cells (gr. *Konios*: dust) are very small.

As can be seen in Fig. 1, structural differences between the three pathways emerge at the level of the retina as each pathway receives projections

from a different group of retinal ganglion cells. The M pathway is based on inputs from diffuse bipolar cells and parasol ganglion cells. Receptive fields of these cells combine inputs from long (L)- and middle-wavelength (M) cones additively (L + M; the achromatic dimension). Although retinal ganglion cells that ultimately project to the P layers are heterogeneous, midwavelength bipolar and midwavelength ganglion cells form the bulk of the P pathway. The receptive fields of P cells which receive inputs from midwavelength ganglion cells oppose inputs from L and M cones (L–M; reddish-greenish dimension). The two intermediate K layers receive projections from small bistratified cells. The receptive fields of such K cells oppose the excitatory input from S cones with the inhibitory input from L and M cones (S–(L + M); bluish-yellowish dimension). K layers also receive inputs from several other cell types which oppose S-cone input to some combination of L and M cone input (+S – ML large bistratified, –S + ML large sparse, and two –S + ML giant sparse ganglion cells), but their role in color vision is still not fully understood [1, 3]. The distinction between the processing of luminance signals



Magno-, Parvo-, and Koniocellular Pathways, Fig. 1 Structure of the lateral geniculate nucleus, including a map of retinal inputs and cortical outputs. There are six thick layers, with M cells found in the lowest two and P cells found in the top four of these layers. K cells form six thin layers. The first K layer is positioned ventral to the first M layer, while the others are intermediate to the six LGN layers. The M cell axons terminate primarily in layer 4Cα of the primary visual cortex (V1) while the P cell axons

terminate primarily in layers 4Cβ and 4A of V1. The six K layers are further subdivided on the basis of their projections. The two dorsal K layers project into layer 1 of the primary visual cortex. The two intermediate K layers receive information from small bistratified retinal ganglion cells as well as several other types of cells with S-cone input and project it into layers 3 and 4A of the V1. The two ventral K layers project into the superior colliculus

(L + M) and the processing of two types of chromatic signals (L–M and S-(L + M)) by geniculate cells had led to their conceptualization as the three orthogonal visual mechanisms. Such a conceptualization is implemented in the DKL space, a physiological color space based on the seminal study of macaque LGN cell response properties conducted by Derrington, Krauskopf, and Lennie [4].

Historical Overview

Based on his observations of three different types of axons in the rabbit optic nerve, Bishop speculated in 1933 that visual information is likely to be processed along morphologically distinct pathways that operate in parallel [5]. An in-depth historical overview of research into parallel visual pathways is provided by Casagrande and Xu [6]. In brief, by the end of the 1980s, pathway-tracing studies on primates confirmed Bishop's speculation, establishing that different retinal ganglion cell classes projected to separate layers in the LGN, which subsequently projected to separate cortical layers. Numerous physiological studies uncovered marked differences in spatial, temporal, and contrast sensitivities between various cell types that were found in the retina, LGN, and the primary visual cortex. Early studies of primate LGN cell properties attempted to relate their findings to extensive research performed on cats' retinal ganglion cells (X, Y, and W types) since the 1950s. Attempts were also made to relate the newly found geniculate visual pathways to the ventral or "*what*" cortical stream, which was proposed to subserve the processing of objects' visual attributes, and the dorsal or "*where*" cortical stream, which was proposed to subserve the processing of objects' locations [7]. In 1988, Livingstone and Wiesel proposed an integrative model of retinogeniculocortical pathways [8], linking the M pathway to the processing of motion due to the transient responses and high achromatic sensitivity of its neurons and the P pathway to the processing of form and color due to its neurons' sustained response to chromatic contrast and weak response to achromatic contrast.

Livingstone and Hubel's model epitomized the modular approach that was prominent at the time, relating discrete brain structures to specific cognitive or perceptual functions. The koniocellular (K) pathway was ignored by Livingstone and Hubel's model. At that time, the K layers in primates were considered to be too thin to contribute substantially to any single cortical module. Between the 1950s and 1970s, the K pathway was generally studied in bush babies and lorises, which are nocturnal prosimian primates [2]. Subsequent studies confirmed that K cells in these species were quite similar to those in other primates. Since then, a range of studies was performed on the primate K pathway, revealing diverse functional properties of K cells, including their role in color vision. At the same time, much more has become known of retinal processing, with 20 different types of retinal ganglion cells identified [9]. Cortical projections of M, P, and K cells have also been mapped more extensively, as will be discussed in the next section.

Functional Considerations

Numerous neuroanatomical and neurophysiological studies have investigated the properties of the three parallel pathways. As noted in the historical overview, continuous attempts have been made to relate the pathways to the processing of different visual properties: motion processing, form processing, luminance and color processing, and so on. As evidence accumulated it became more and more clear that simple and direct associations between physiological responses and perceptual properties are difficult to establish [1, 10, 11]. In fact, M, P, and K pathways can only tentatively be treated as wholly separable when it comes to their function: for example, the M pathway contains cells that project information into dorsal cortical areas subserving motion processing, which fits Livingstone and Hubel's model, but it also contains cells that project information into ventral areas subserving object representation [6].

The most robust and the least disputed evidence for differential processing between M, P,

and K pathways comes from primate color vision research. Strong links have been established between the M pathway and the luminance (L + M) channel. The chromatic L–M channel, which computes the difference between L and M cone signals providing a reddish-greenish dimension of color, maps on to the cells of the P pathway. Furthermore, the L–M cone-opponent mechanism provides signals that could be transmitted in combination with luminance information [4]. The view that such “multiplexing” through the L–M channel underlies achromatic vision is based on the fact that P cells are much more numerous than M cells and was initially strengthened by observations of poor spatial acuity of M cells, although recent data showed much smaller differences in M and P cell receptive field center sizes than originally found. In fact, there is an opposing view that strict subcortical segregation of luminance and L–M signals would be beneficial, since it would optimize signal transmission in a noisy, bandwidth-limited system such as primate vision [1].

Unlike P cells, K cells that receive S-cone input from small bistratified cells are not considered to make a significant contribution to luminance processing. Their receptive fields are larger than those of luminance-sensitive cells [12], compromising their usefulness for high-acuity spatial vision. Their signal is amplified cortically, being weakly present in many cells from V1 onward [13]. S-cone-sensitive K cells may still contribute to functions other than just color appearance. For example, they project directly to motion-sensitive cortical area (MT/V5) [10].

While it is not possible to fully equate geniculate pathways with psychophysically identified luminance and chromatic channels, a certain degree of functional overlap between them is evident and has been used in many studies that rely on luminance or chromatic contrast to attempt to isolate or bias visual processing toward different pathways. As M pathway dysfunctions have been hypothesized to underlie schizophrenia, autism spectrum disorders, dyscalculia, and dyslexia, the ability to isolate this pathway through carefully designed stimuli is of keen interest to clinical neuroscience. Studies in this field do not only

compare responses to luminance stimuli with those to isoluminant stimuli but also often rely on other well-known processing differences between parallel visual pathways. Fast temporal frequency, low luminance contrast, and low spatial frequency are generally thought to bias processing toward the M pathway, while low temporal frequency, high chromatic contrast, and high spatial frequency are assumed to bias processing toward the P pathway. It is paramount to be careful when interpreting the findings of such studies. For example, besides many transient response neurons, the M pathway also contains some neurons with sustained responses. Similarly, while contrast sensitivity functions for luminance and chromatic stimuli do indeed differ, many spatial frequencies that lead to larger excitation in one pathway happen to concurrently cause some excitation in the other pathway. Therefore, the level of bias to one or the other pathway is very difficult to quantify. Finally, responsiveness of M, P, and K pathways to various visual features cannot be assumed to translate directly into perceptual sensitivity for these features. For example, P cells with receptive fields in the parafovea are sensitive to a wide range of temporal frequencies, but human observer performance reveals that the sensitivity of the visual system as a whole does not match the sensitivity that is present in this geniculate circuitry [14]. Thus, some of the sensitivity evident in P cells is not utilized at subsequent stages of processing that drive behavior.

Isolating subcortical processing streams or biasing processing toward a certain pathway is ultimately complicated by the fact that the separation of signals from different visual pathways is not maintained in the cortex. Receptive fields in V1 are larger than in the LGN, indicating that each V1 cell receives input from several LGN cells. S-cone signals become more prominent in the cortex, with weak S-cone signals being found in many V1 neurons. Finally, LGN cells with chromatic sensitivity show preferential responses to stimuli that align with the S-(L + M) and L–M mechanisms, but V1 cells have much more widely distributed color preferences. This indicates that S-(L + M), L–M, and L + M signals start interacting in the cortex. In fact, most of the

early visual cortex contains neurons that are tuned to both color and luminance while also being sensitive to certain spatial properties, e.g., orientation [13].

Summary

Magno-, parvo-, and koniocellular pathways are established at the level of the LGN. They receive inputs from different retinal ganglion cells and project to different layers of the V1, leading to parallel processing of visual information. Parallel visual pathways have been extensively researched both in terms of their structure and in terms of their functional properties. Geniculate pathways are of interest for color scientists as the retinal projections they receive differ in chromoluminance content: while the M pathway is solely devoted to the processing of luminance information, the P and K pathways also subserve chromatic processing. Other differences between pathways concern contrast sensitivity and spatial and temporal processing, with the M pathway being preferentially responsive to high temporal but low spatial frequencies and to low luminance contrast. Despite these recognized differences, studies that attempted to relate these pathways to specific visual functions have met with limited success. Such research is complicated by the fact that data on physiological responses of M, P, and K cells does not always correspond to behavioral data obtained from human observers under the same conditions [14]. This is at least partly caused by the fact that visual signals undergo much further processing in the cortex, as signals from different pathways get recombined from primary visual cortex onward [11]. There is growing evidence that contributions of M, P, and K pathways to various perceptual functions are not as segregated as was once thought. Recent research has also revealed that subcortically there are likely to be more than just three parallel visual pathways, based on the evidence that the processing in the retina itself is already very complex and driven by a wide array of cells with different inputs and connectivity [15].

Cross-References

► [Ganglion Cells](#)

References

1. Lee, B.B., Silveira, L.C.L.: Cone opponency: an efficient way of transmitting chromatic information. In: Kremers, J., Baraas, R.C., Marshall, N.J. (eds.) *Human Color Vision*. Springer, Char (2016)
2. Hendry, S.H.C., Reid, R.C.: The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* **23**, 127–153 (2000)
3. Kremers, J., Silveira, L.C.L., Parry, N.R.A., McKeefry, D.J.: The retinal processing of photoreceptor signals. In: Kremers, J., Baraas, R.C., Marshall, N.J. (eds.) *Human Color Vision*. Springer, Char (2016)
4. Stockman, A., Brainard, D.H.: Color vision mechanisms. In: Bass, M. (ed.) *OSA Handbook of Optics*, 3rd edn, pp. 11.11–11.104. McGraw-Hill, New York (2010)
5. Bishop, G.H.: Fiber groups in the optic nerves. *Am. J. Phys.* **106**, 460–470 (1933)
6. Casagrande, V.A., Xu, X.: Parallel visual pathways: a comparative perspective. In: Werner, J.S., Chalupa, L.S. (eds.) *The New Visual Neurosciences*, pp. 494–506. MIT Press, Cambridge (2004)
7. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Goodale, M.A., Mansfield, R.J.W. (eds.) *Analysis of Visual Behavior*, pp. 549–586. MIT Press, Cambridge (1982)
8. Livingstone, M.S., Hubel, D.H.: Segregation of form, color, movement and depth: anatomy, physiology and perception. *Science*. **240**, 740–749 (1988)
9. Conway, B.R., Chatterjee, S., Field, G.D., Horwitz, G. D., Johnson, E.N., Koida, K., Mancuso, K.: Advances in color science: from retina to behavior. *J. Neurosci.* **30**(45), 14955–14963 (2010). <https://doi.org/10.1523/jneurosci.4348-10.2010>
10. Kaplan, E.: The M, P and K pathways of the primate visual system revisited. In: Werner, J.S., Chalupa, L.S. (eds.) *The New Visual Neurosciences*. MIT Press, Cambridge (2012)
11. Johnson, E.N., Mullen, K.T.: Color in the cortex. In: Kremers, J., Baraas, R.C., Marshall, N.J. (eds.) *Human Color Vision*. Springer, Char (2016)
12. Buzás, P., Kóbor, P., Petykó, Z., Telkes, I., Martin, P. R., Lénárd, L.: Receptive Field properties of color opponent neurons in the cat lateral geniculate nucleus. *J. Neurosci.* **33**(4), 1451–1461 (2013). <https://doi.org/10.1523/jneurosci.2844-12.2013>
13. Solomon, S.G., Lennie, P.: The machinery of colour vision. *Nat. Rev. Neurosci.* **8**(4), 276–286 (2007). <https://doi.org/10.1038/nrn2094>

14. Lee, B.B.: Visual pathways and psychophysical channels in the primate. *J. Physiol.-Lond.* **589**(1), 41–47 (2011). <https://doi.org/10.1113/jphysiol.2010.192658>
15. Rider, A.T., Henning, G.B., Eskew, R.T., Stockman, A.: Harmonics added to a flickering light can upset the balance between ON and OFF pathways to produce illusory colors. *Proc. Natl. Acad. Sci.* **115**(17), E4081–E4090 (2018). <https://doi.org/10.1073/pnas.1717356115>

Mass Pigmenting

- [Coloration, Textile](#)

Material Origins of Advanced Color Terminology

- [Materials and the Origins of Abstract Color Naming](#)

Materiality, Luxury, Hue, Saturation, Brightness, Intensity Linguistically Expressed in Ancient Languages

- [Evolution of Color Expression](#)

Materials and the Origins of Abstract Color Naming

David Alan Warburton
Institute for the History of Ancient Civilizations,
Northeast Normal University, Changchun, China

Synonyms

[Material origins of advanced color terminology;](#)
[Precious stones and metals at origins of linguistic expression of color](#)

Definition

Materials – integrating hue, brightness, intensity, hardness, and occasionally form, into a single category, e.g., jade (axes), gold (beads) – were decisive for the development of increasingly abstract linguistic color categories from the era of the earliest documented language sources.

Material Color

Extremely controversial is the linguistic importance of materiality, with colors of materials used in personal adornment and ornamentation contributing to the creation of abstract color terminology. In the Berlin & Kay approach, abstraction involved the denial of materiality and materiality was thus excluded; only abstract linguistic evolution was aimed at.

However, even in the era of the wide-spread application of the Berlin & Kay Theory, in the treatment of the earliest languages as evolutionary stages, materiality was unavoidably accepted. Thus, while philologically included, it was theoretically denied in the case of Ancient Egyptian “white” = “silver” and “green” = a mineral or stone, as well as with Sumerian “white” = “silver” and Akkadian “red” = “carnelian” [5]. Nevertheless – since precisely these concepts are fundamental to the development of lexical color – materiality is increasingly playing a role in contemporary theoretical discussions about ancient color (e.g., [1–4, 6]). Significantly, the story goes back virtually to the origins of writing. (see also Encyclopedia of Color Science and Technology entries on ► [“Evolution of Color Expression”](#) ► [“Ancient Color Terminology”](#)).

The problem is understanding the cognitive relationship between perception and expression as actually experienced in language. For ages, mankind has seen the sun shining in the sky. Yet, in the second millennium BC some Egyptian poets described the visual experience of the evening sun shining on the Nile as “inundating [it] with gold” (Fig. 2) or the morning sun itself “rising golden” in a “field of turquoise” ([6]). Such expressions are rare and seem poetic, but



Materials and the Origins of Abstract Color Naming, Fig. 1 Statue of a standing ram (h: ca. 50 cm) from the mid-third millennium BC elite cemetery at the Mesopotamian city of Ur. Made with gold, silver, copper alloy, lapis lazuli, shell, red limestone, bitumen ([11])

the ancient Egyptians could not express it differently, for they had no independent words for “yellow” or “light blue.” Thus, in a flight of poetic fancy, someone used the precious materials with which the Near Eastern elite had become familiar (Figs. 1, 2, 3, and 4) to describe something that mankind had experienced daily for ages – but had never been put into linguistic expression using abstract words for color, because there were none. And this completely new – but unfinished – solution would endure for another millennium. In Egyptian, the first references to the “sky” as having a color itself are Greco-Roman. By contrast, the first reference to sun as being golden lies in the first half of the third millennium BC.

Earlier, when defined, the sky was turquoise or lapis lazuli (rather than having a color itself) – and this usage only ended during the first centuries AD. Lapis lazuli and turquoise were treated as



Materials and the Origins of Abstract Color Naming, Fig. 2 The Nile in the evening with the sun inundating the river with gold, as expressed by an ancient Egyptian poet. Note that the sky too is golden; the concept of sky and water being “blue” only came after millennia of confrontations with the irregularities of reality and the development of linguistic expression, which contributed to the gradual emergence of abstraction whereby water and sky “became (normally) blue”

precious stones; from the late third millennium onward, gold and lapis lazuli had prices higher than silver (which served as money). The statues of the gods in the temples were made of granite, but also such of materials, some of which were the origins of what later became color words, some merely served to identify “brightness,” and Plato considered “bright” to be a color.

Sumerian “white,” “bright,” and “pure” were linked to silver *kù-babbar*, which was the “(pure) white metal” and was then understood as “white” (*pešû*) in Akkadian. In earlier Egyptian (*ḥdj*) was the word for silver and “white.” In Greek, silver played a role in “shining” (*argos*). Other materials, such as jade (Fig. 4) may have influenced the



Materials and the Origins of Abstract Color Naming,

Fig. 3 Silver hair pin (h: ca. 25 cm) from the mid-third millennium BC elite cemetery at Mesopotamian Ur; silver, lapis lazuli, carnelian, gold, shell, red limestone ([11]). Silver tarnishes and thus “brightness” is lacking today

choices without leaving many traces. Although they did not invariably lead to color names, the materials played a role in the formation of the understanding of color, as can be seen in the earliest languages.

These particular types of materials were largely used only by humans during the last 10,000 years (Bar-Yosef Mayer, Moutsiou, Warburton in [2]), and in the earliest languages, they were rarely used as color words – to the degree that Schenkel ([7, 8] and in [2]) probably correctly disputes that these words played an important cognitive role in the ancient Egyptian expression of color – regardless of developments in other languages later.

Schenkel thus dismisses their importance, whereas Warburton uses exactly these as prominent cases of words, which became color terms when the materials and related words wandered. The word “cyan,” a “blue” in English, owes its origins to Classical Greek *kuáneos*, of which the Modern Greek variant *kuano* is still used for

“blue.” A process of linguistic exchange gave birth to the abstract color term *kuáneos*, but only in Classical Greek where *kuáneos* means “dark blue” or “dark.” Derived from *ku-wa-no*, which in Mycenaean Greek Linear B meant “(dark blue) glass (paste)” (Fig. 6), the word *ku-wa-no* itself was derived (via Ugaritic and/or Hittite) from *uqnû*, an Akkadian word, which was itself originally a loan coming from some other, older, unknown language where it was the name of the mineral lapis lazuli (which was the meaning in Akkadian). In Sumerian, Egyptian, and Akkadian different terms for the name of the stone were also used for glass and dyes of a blue color. The French *azur* (*la couleur du ciel*) likewise owes its origins to lapis lazuli, but through the Persian word *lâzaward* ([14]).

Likewise, relatively late in the history of the development of color terminology, in Classical Greek, *xrusós* (literally “gold”) and words related to it (e.g., *xrúseos*) can mean “golden” or “yellow”; its predecessor in Mycenaean Linear B was *ku-ro-so*, which was more closely related to the material. The word itself is related to Akkadian *hurāsu* (literally “gold”), which was itself usually restricted to the material, but things could be “like gold” in color. In Akkadian, the nouns lapis lazuli (*uqnû*, above) and carnelian (*sāmtu*; related to “redness,” “red” in Akkadian) appear regularly together in the same texts, along with *hurāsu* and *hurāšānû*, “gold” and “golden” (as colors and as jewelry – so that both hues and stones are meant). Thus, the Greek was building on earlier usage, and Mycenaean reflects an intermediate phase.

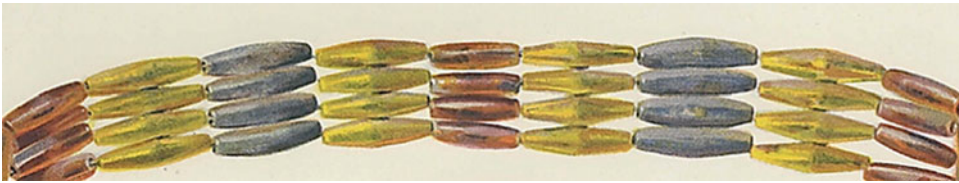
Loan-concepts reveal loanwords, etymologies, and materials. Pronounced in modern Danish, *guld* (“gold”) and *gul* (“yellow”) are virtually identical – and both linked to English “gold” and “yellow” as well as German *Gold* and *gelb*. In several languages therefore, the word for “gold” implies color associations mostly from the “yellow” portions of color space but can also imply “orange” (and even “red”) areas of color space – and also “bright.” Ancient Chinese *jīn* (literally “gold”) did not designate a color so much as “bright, shining” when used figuratively [10]. The Ancient Egyptian *nb.w* is not related to



Materials and the Origins of Abstract Color Naming,

Fig. 4 Idealized typical European Neolithic axe of jadeite (l: ca. 20 cm) dating to the mid-fifth millennium BC; luxury items of Alpine jade were spread from the Atlantic to the Black Sea coast ([13]), where the earliest lozenge-shaped gold beads were found (virtually identical to, but antedating the third-millennium beads of Fig. 5 by some two millennia). Chinese use of green jades paralleled the beginnings in Europe, but did not stop; Mesoamerican usage came much later, and maintained the tradition. Conceptual

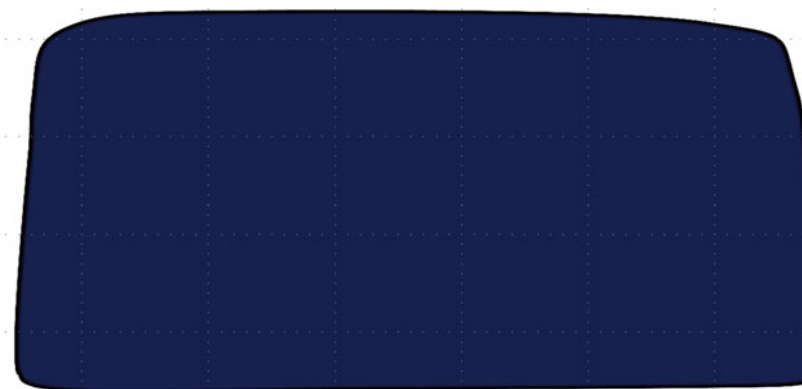
exchange where material, form, and hue were symbolically linked took place in a cultural *koiné* exceeding the spatial and temporal zones generally incorporated into current research: one early greenstone axe was found in Tasmania (Warburton in [2]); one of the earliest is from Neolithic Syria and dates to ca. 10,000 BC [12]. The conceptual traditions binding form, hue, and material antedate the known linguistic phenomena – and therefore transgress the boundaries of known languages, as elaborated in the text



Materials and the Origins of Abstract Color Naming,

Fig. 5 Necklace (each bead l: 2 cm), mid-third millennium BC elite cemetery at the Mesopotamian city of Ur ([11]). The beads visible here are lozenge-shaped, alternating red (carnelian), yellow (gold), and blue (lapis lazuli). A similar set of lozenge-shaped beads – with gold, carnelian, turquoise, and lapis lazuli (i.e., yellow, red, dark blue,

light blue/light green) was found on a bracelet from Egypt one thousand years later (fifteenth century BC). The earliest gold beads of this type were found in Varna on the coast of the Black Sea (4400 BC); a necklace of similar lozenge-shaped beads made of a green stone, dating to the third millennium was found on the Greek island of Naxos



Materials and the Origins of Abstract Color Naming,

Fig. 6 Sketch of a Bronze Age glass ingot (ca. h: 7 cm, Ø 15 cm) from a shipwreck in the eastern Mediterranean, ca. 1300 BC. The glass was probably manufactured in Egypt (where lapis lazuli was called *xsbdfj*), but in the

Mediterranean was probably sold as *uqnû*, the Akkadian word for lapis lazuli, giving rise to Ugaritic *iqnu* and Mycenaean Greek Linear B *ku-wa-no*, meaning “glass.” The word lives on as “cyan”

these Chinese, Semitic, and Indo-European words, but mainly designated the material – yet occasionally the color “yellow” (as noted, above, of the sun) cf. Fig. 2.

For millennia Greek maintained a couple of words for “yellow,” neither of which was distinctly “yellow” – while also using “gold.” The earliest languages thus did not know abstract “yellow” and associated “gold” with the gold (en) metals, but somehow color and material were related and over time led to the linguistic partitioning reflected in, e.g., English. The strong color significance of gold, familiar today, is probably related to the transfer of the word (along with the material) through space and time.

The Akkadian color term *xašmānum* for “blue-green” or “light blue” ([6]) was itself a loanword from Egyptian, designating the light blue or purple stone “amethyst.” Although the Egyptian *ḥsmn* is not a color term, the Egyptian word was adopted as a color word in Akkadian. But the Egyptian was not used as a color word in the Mycenaean world – despite massive imports of the actual stone amethyst from Egypt (Warburton in [2]). Using the Egyptian word was superfluous as other – Levantine – imports gave the Mycenaeans *po-pu-re-jo*, which went into Classical Greek as *porphúreos* and survives in “purple” today. That the Greeks understood the color as being more red than blue is clear because the Romans treasured the stone porphyry. Both amethyst and porphyry are found in Egypt, and thus it is curious that the original – Egyptian *ḥsmn* amethyst purple – hue associated with the word in English today seems to have survived the mutations, despite the fact that the reddish hue associated with the Greek term *porphúreos* – which is etymologically the origin of the modern word – was different.

Although vegetation is known to be “green,” the Egyptian *wādj* (green) is frequently associated with a classifier signifying a stone. Although the Egyptian *wādj* is related to the Akkadian *warqu*, the category centroid, or focus, of the Egyptian term was in “green” and neither “green-yellow” nor “green-blue” – and not “yellow” (Schenkel in [2, 8]). By contrast, the Akkadian *warqu* denoted both “green” and “yellow” appearances and was

used to describe both vegetation and gold; it definitely did not mean “green-blue” or “green-yellow”. However, while “green” was – with a loanword – linguistically demarcated in Akkadian and Egyptian, and the Egyptian linked to a stone; dark green stones are known, but not common in Ancient Near Eastern jewelry.

It is therefore odd that the Akkadian and Egyptian are related in a fashion which suggests both can be traced back to an older word, and that the Egyptian is linked to a stone. Neither of the terms is a loanword from the other language, the Egyptian having changed the /r/ → /ā/ before the literate era. The modern European terms “green,” *grün*, and *grøn* share parts of the root as do *viridis*, *vert*, and *verde*. Thus, in *grün* and *vert*, contemporary German, English, Italian, and French bear traces of different elements, which probably go back to before third millennium BC Akkadian *warqu* and Egyptian *wādj*, and are related to them. All these would therefore apparently go back to an older Pre-Indo-European word (*√qrw or *√wrq?) in an unknown language, antedating the invention of writing and never recorded. Unrelated Sumerian *gin*, likewise inherited from Prehistory, may have meant “axe.” Curiously, when the Indo-Europeans arrived in Europe, the earlier population was dispersed and the practice of using ritual axes made of jade ceased ca. 3000 BC. Yet the original Neolithic stones were familiar before then, and the original word may therefore be derived from a Pre-Indo-European European Neolithic designation for “jade,” “axe,” or “greenstone” (Fig. 4).

When the celebrated Levantine organic “purple” dye (the organic product leading to the modern word, but producing a “red”) entered the Mycenaean world, the Mycenaeans also apparently adopted the custom of manufacturing the dye themselves, but continued to use the loanword. In the case of the famed purple dye, the cost was apparently so high that it was understood as a luxury itself. By contrast, when the Egyptians and Babylonians manufactured glass themselves, they used foreign loanwords meaning “lapis lazuli” (*xsbaj* and *uqnū* respectively) to designate the glass they themselves produced; Thavapalan makes the same point for calcite and yellow glass

([2, 3]). Thavapalan stresses brightness, while Warburton favors recognizing the contrasting hues distinguishing gold, silver, lapis lazuli, carnelian, turquoise, and jade as the reason that the names of the stones were applied to glass, faience, and dyes.

An Akkadian “red,” *ruššu*, is probably the same as Greek *eruthrós* and Italian *rosso*. The Mycenaean Greek Linear B *e-ru-to-ro* is doubtless the same word as Classical Greek *eruthrós* – and will be related to *ward*, the word for “rose” in Arabic and *vr̥dhi*, a similar term in Sanskrit. It is possible that *rwdj* was a red in ancient Egyptian, thus related to Akkadian, Greek, Italian, Arabic, and Sanskrit – but Schenkel disputes this (see in [2, 7, 8]; also [6]). Yet, even if details require modification, the principle is clearly that the ancient concepts moved between languages.

Beyond that are the colorful materials used by the Mayas, Chinese, Europeans, Egyptians, and Mesopotamians that did not invariably enter into color terminology, but played a major cognitive role in art: ivory, shell, obsidian, black granite, red limestone, and agate are among those that do not figure prominently (or at all) as color terms. Other such materials which were used prominently, but did enter into terminology were ocher, carnelian, gold, and amethyst.

There is an enormous difference between (a) the use of precious materials as the foundation of a color terminology and (b) the role of color in the craftsmanship of manufacturing art, where words appear for pigments, but did not become “color-terms.” In the earliest documented languages, in Egypt and the Near East, inorganic materials were largely used for producing pigments, for practical reasons (and the hues linked to the more precious materials being imitated, e.g., Blom-Böer in [2]) while in the later Mesoamerican tradition inorganic pigments were used on stone, but organic pigments were employed when applying color to organic materials (as opposed to stone), with strong symbolic connotations ([4], Dupey-García in [1]). By the first millennium BC, dyes with organic components had become more important in the Near Eastern

world, but the formative vocabulary had been determined by the inorganic materials.

Precious Materials, Loanwords, Abstraction, and Grammar

Language interaction was important in trade as the valuable materials were sought out, leaving traces in disparate ways, where forms (e.g., axes and lozenge-shaped beads) also played a part. Thus, jade, carnelian, silver, lapis lazuli, and gold all played roles in the development of the use of color. Via art, these filtered into elite consciousness and the earliest known languages, with perceptions and expressions feeding into the development of color terminology. Materials eventually led to abstract color words, but usually only in the languages into which they were imported, a process beginning in the third millennium BC, but only having systematic effects from the first millennium BC onward. Through the second millennium BC, many expressions of color are rooted in materials – most of which were later eclipsed with abstract words created and adopted through exchange.

One of the greatest obstacles to understanding the nature of the earliest origins of ancient color terminology (in the Ancient Near East) and the origins of abstraction (in Greek and Chinese) is appreciating that in the earliest usage, the ancients did not classify the world according to modern terminological divisions of the diversity of visible light, but rather that they initially used their perceptions of precious materials to express many of the colors they perceived. Thus, their understanding of color was rooted (in the sense of hue, saturation, brightness, intensity, hardness, and frequently form) in the materials – which properties they then applied to other domains. This eventually created the basis for abstract color terminology. However, the origins render the discussion complicated. (See, e.g., [1–3, 6, 9, 13]).

In Akkadian, Chinese, and Greek, color words are largely adjectives. Some of the Greek terms can be traced back to Near Eastern materials. Of the languages discussed here, Chinese is the only

language with a completely abstract color vocabulary, where hues and terms match, more or less. Although silk appears as a component in the writing of some color words in Chinese, even as a component, in China jade played an even more marginal role. Thus Chinese color terms never bore close primary relations to precious, natural object categories; the evolution of Chinese color vocabulary differed fundamentally from that of the West – but the results are similar to those in Greek (probably implying diffusion rather than convergence).

The Mediterranean languages usually had a word for “color” as a phenomenal experience; however, the words were not restricted to a single meaning in terms of hue. In Egyptian, the word *jwn* is probably derived from the designation for a “vein” of ore (meaning the material color which was visible), but it also meant “character” of a person in the figurative sense (what was hidden under the surface, but confusingly also skin itself). Although in later Chinese the suffix *sè* is frequently attached to color words, in the earliest Chinese no term for color as such has been discovered.

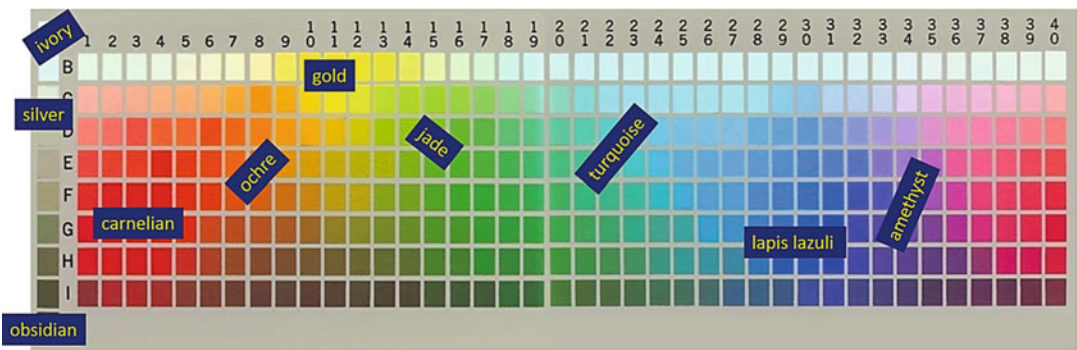
In his list of Indo-European etymologies, Shields ([15]) suggests that the abstract terms “bright,” “gleaming,” etc., are the basis for many words that eventually became abstract color words. However, Chinese terms for “bright” or “clear” such as *míng* and *qīng* did not give rise

to color words. Over time, Classical Greek *glaukós*, initially denoting “gleaming,” did acquire a meaning of “bluish green” or “gray” – but never became a stable color term. Sumerian *šuba* means “shining” but never became a color term; *tjhn.t*, an Egyptian word for “gleaming” or “dazzling,” never became a color word. In ancient languages, there are many more words for “gleaming,” “shining,” “brilliant,” “bright,” etc., than can be inventoried (see also Encyclopedia of Color Science and Technology entries on ► “Evolution of Color Expression” and ► “Ancient Color Terminology”).

Many of these terms, like *tjhn.t*, *hsmn*, and *šuba* probably originally designated precious stones, and the origins of “bright” may well lie in this domain (rather than abstract categories). Thus, through diffusion, precious materials (rather than abstract “brightness”) are more likely the origins of both “brightness” and abstract color words. Gold, silver, carnelian, turquoise, and lapis lazuli suggest that, via hues, a select few materials bridged the way into abstraction as color terms; others served a similar role in the “bright” and “dark” categories. Along with hardness, form, brightness, and intensity, hue was probably decisive in determining the names and colors that were developed or maintained; abstract “bright” is not evident.

Figure 7 approximates the relationships between color appearances and color terms.

M



Materials and the Origins of Abstract Color Naming, Fig. 7 Examples of physical materials related to ancient color terminology displayed in congruent regions of a Munsell Color Chart

Summary

Through art and language, precious materials played a decisive role in the early understanding of color. Appreciation of a wider spectrum of colors (than the dark/bright/red distinction which must be more than 10,000 years old) is a product of social developments in the last 10,000 years – and mostly in the last 5000 years. A tradition of luxurious artwork began in the Ancient Near East, and brought color into the lives of the members of the elite, at a time when book-keepers counting stones could also be poets. Over time, the manufacture of artificial pigments used as textile dyes and in the production of glass brought color to increasing proportions of the population. Before this, the trade in semi-precious stones and precious metals contributed to the development of abstraction as the materials changed hands, frequently bearing the earlier name as a loanword, which was incorporated into the language of the recipients – and used to designate both the artificial materials and the colors in a fashion which had not hitherto been normal. In this fashion, materials contributed to the emergence of abstract color terminology – but in the languages of the recipients (initially certainly Chinese and Greek) and their successors, such as English and German, rather than in the languages where the earliest terminology developed (in Egypt and Mesopotamia).

Cross-References

- [Berlin and Kay Theory](#)
- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Effects of Color Terms on Color Perception and Cognition](#)

References

1. Goldman, R. (ed.): *Essays in Global Color History: Interpreting the Ancient Spectrum*. Gorgias Press, Piscataway (2016)
2. Thavapalan, S., Warburton, D.A. (eds.): *The Value of Color: Material and Economic Aspects in the Ancient*

World. Berlin Studies of the Ancient World 70, Berlin (2019). <https://doi.org/10.17171/3-70>

3. Thavapalan, S.: *The Meaning of Color in Ancient Mesopotamia. Culture & History of the Ancient Near East 104*. Brill, Leiden (2019)
4. Houston, S.D., Brittenham, C., Mesick, C., Tokovinine, A., Wariner, C.: *Veiled Brightness: A History of Ancient Maya Color*. University of Texas Press, Austin (2009)
5. Baines, J.: Color terminology and color classification: ancient Egyptian color terminology and polychromy. *Am. Anthropol.* **87**, 240–262 (1985)
6. Warburton, D.A.: The theoretical implications of ancient Egyptian colour vocabulary for anthropological and cognitive theory. *Lingua Aegyptia.* **16**, 213–259 (2008)
7. Schenkel, W.: Die Farben in Ägyptischer Kunst und Sprache. *Z. Ägyptische Sprache.* **88**, 131–147 (1963)
8. Schenkel, W.: Color terms in ancient Egyptian and Coptic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 211–228. Benjamins, Amsterdam (2007)
9. Warburton, D.A.: Colors in bronze age Egyptian art and language. In: Brinkmann, V., Primavesi, O., Hollein, M. (eds.) *Circumlitio: the Polychromy of Antique and Mediaeval Sculpture*, pp. 170–187. Liebighaus, Frankfurt (2010)
10. Schuessler, A.: *ABC Etymological Dictionary of Old Chinese*. University of Hawai'i press, Honolulu (2007)
11. Woolley, C.L.: *Ur Excavations. Volume II: the Royal Cemetery. Plates*. British Museum and Museum of the University of Pennsylvania, Oxford (1934)
12. Stordeur, D.: *Le Village de Jerf el-Ahmar*. CNRS Editions, Paris (2015)
13. Pétrequin, P., et al.: *Jade: Grandes Haches Alpines du Néolithique européen, Ve et IVe millénaires Av. J.-C.* Presses Universitaires de Franche-Comté, Besançon (2012)
14. Robert, P.: In: Rey-Debove, J., Rey, A. (eds.) *Le Nouveau Petit Robert. Le Robert*, Paris (2007)
15. Shields, K.: Indo-European basic colour terms. *Can. J. Linguist.* **24**, 142–146 (1978)

Matrix Beam

- [Automotive Front Lighting System \(Status 2020\)](#)

Matting

- [Compositing and Chroma Keying](#)

Mauveine

Harold S. Freeman

Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Definition

Mauveine, also known as Perkin's violet, mauve, and aniline purple, is generally acknowledged to be the first synthetic dye, having been discovered by William Henry Perkin in 1856 [1]. As shown in Fig. 1 below, Mauveine is a cationic dye having an acetate or sulfate moiety as a typical counterion. It is also an azine dye belonging to the Safranine class and was derived from oxidation of aniline in combination with a mixture of *ortho*- and *para*-toluidine [2]. Interestingly, Mauveine was found to give on silk fabric a regal purple color that was comparable to the shade produced by the natural dye Tyrian purple. It was the dominant synthetic colorant in the marketplace until 1873, giving way to triarylmethane dye fuchsine (rosaniline), which was brighter and more easily made.

Historical

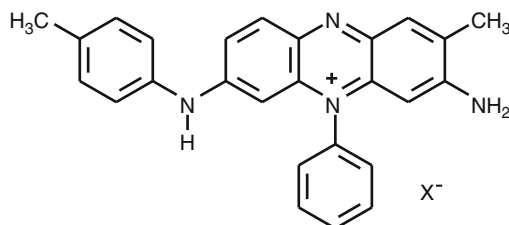
The synthesis of Mauveine began in 1856 when the 18-year-old William Henry Perkin attempted to produce the antimalarial agent quinine ($C_{20}H_{24}N_2O_2$) from *N*-allyl-*para*-toluidine ($C_{10}H_{13}N$). Bearing in mind the molecular formulas of the starting and target compounds but

without a knowledge of their molecular structures, it was anticipated that quinine was an oxidative dimerization step away from the starting compound. The brown powder obtained as the product of this reaction lacked any quinine. Likewise, Perkin did not produce quinine following oxidation of the parent compound aniline using potassium dichromate. However, the black solid obtained gave a purple solution when hot ethanol was used to help clean the glassware employed in this synthesis [3]. Subsequent analysis of this purple product revealed it could be useful as a dye for silk and other textile fibers [4]. With this in mind, Perkin patented his new technology and in 1857 he and his family established a company (Perkin and Sons) to produce the dye on an industrial scale. The dye was originally called aniline purple. Two years later, their product was called mauve in England, from the French name for the mallow flower, and chemists subsequently referred to it as Mauveine [5]. During 1859–1861, Mauveine was the leading dye in the fashion industry.

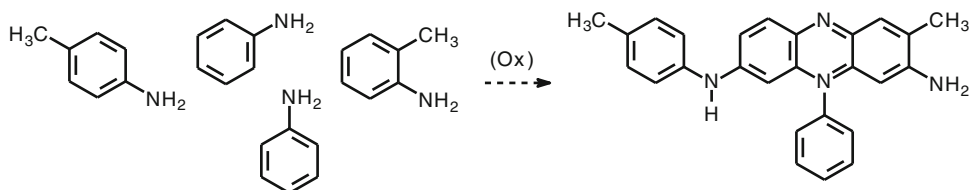
Dye Formation

The origin of Mauveine was an act of serendipity in two respects: 1) an entirely different but highly useful product was the outcome of the experiment conducted and 2) pure aniline was not used, as it would not have produced Perkin's dye. As the structure of Mauveine indicates, the formation of Mauveine occurred because nineteenth-century aniline contained significant amounts of *ortho*- and *para*-toluidine impurities, which is illustrated in Fig. 2. Commercially, the aniline mixture is dissolved in sulfuric acid and water in approximately a 1:1:2 ratio, then potassium dichromate is added. In the Fig. 2 illustration, the Mauveine structure is generated from two equivalents of aniline and one equivalent each of *ortho*- and *para*-toluidine.

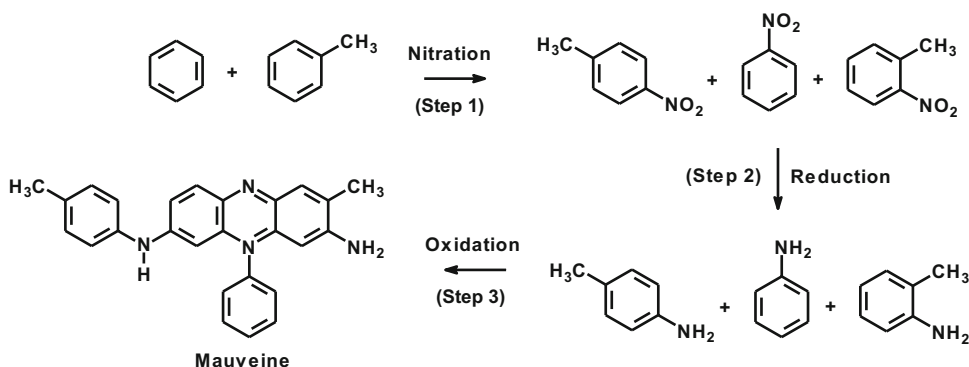
In the nineteenth century, the Perkin and Sons production of Mauveine was actually a three-step process, beginning with commercially available benzene, as shown in Fig. 3. The goal of steps 1 and 2 was to produce large quantities



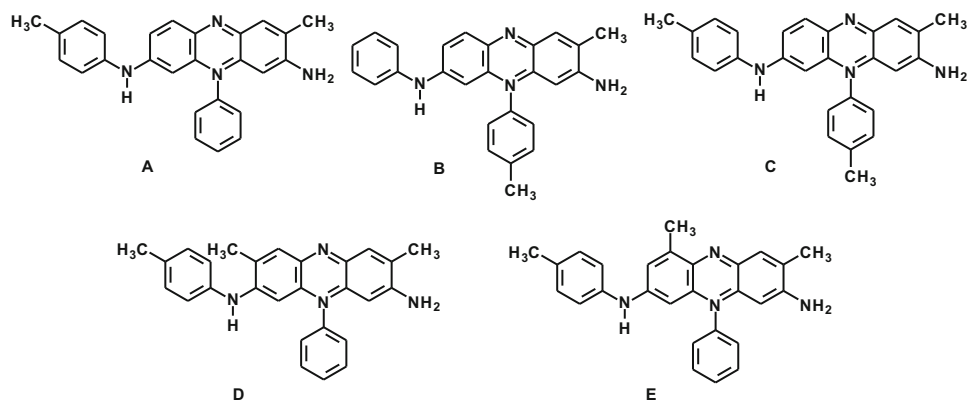
Mauveine, Fig. 1 Molecular structure of Mauveine, where X = acetate, sulfate



Mauveine, Fig. 2 Illustration of Mauveine formation from mixed anilines



Mauveine, Fig. 3 Steps required to obtain aromatic amines used in Mauveine manufacture



Mauveine, Fig. 4 Examples of structures reported for Mauveine

of the anilines necessary for dye manufacture. At that time benzene contained a significant amount of toluene, which gave a mixture of nitrobenzene and isomeric nitrotoluenes. Reduction of the nitro compounds gave the required aniline mixture. Mauveine formed when the mixture of aromatic amine was treated with $K_2Cr_2O_7/H_2SO_4$.

Structures

In view of the possible combinations that can occur when aniline, ortho-toluidine, and para-toluidine react with each other in the oxidation step, several “Mauveine” structures have been reported. Examples are **A–E** shown in Fig. 4, which include products containing aniline or

para-toluidine built into the central ring and/or attached to an external ring. Until the work of Meth-Cohn and Smith [6], it was often reported that structure **D** corresponded to the main component of Perkin's Mauveine. However, the 1994 isolation of the principal components of an original dye sample followed by spectroscopic analyses revealed dye **A** to be the major component, with dye **E** the secondary component. These results are consistent with Perkin's recognition of the importance of *ortho*- and *para*-toluidine in the color and properties of Mauveine. Meth-Cohn and Smith also pointed out that structure **D** would require *meta*-toluidine as a key precursor, despite its very minor contribution (4%) to the product mixture following the nitration and subsequent reduction of nitrotoluenes. Compared to dye structure **A**, structure **E** contains an additional *ortho*-toluidine moiety, which is consistent with its significant presence in the anilines mixture used in dye manufacturing. A method for reproducing the synthesis of Perkin's commercial Mauveine has been reported [7]. The yield was quite modest but sufficient to validate the method.

Further Comments

Mauveine is extraordinary not only as the first synthetic and commercial dye but also for its overall contribution to the field of organic chemistry. This important discovery gave birth to the chemical industry and subsequently the pharmaceutical industry, in that the technology developed to manufacture this dye on an industrial scale proved invaluable in this broader context. Similarly, the recognition of its ability to dye silk, and later other fiber types, had a profound impact on the nineteenth-century textile industry. The invention of no other synthetic dye has been more impactful.

References

1. Hunger, K.: Dyes, General survey. In: Hunger, K. (ed.) Industrial Dyes: Chemistry, Properties, Applications,

- p. 1. Wiley-VCH Publishers, Weinheim, Germany (2003)
2. Allen, R.L.M.: Colour Chemistry, p. 124. Thomas Nelson and Sons, Edinburgh, Scotland (1971)
3. Gordon, P.F., Gregory, P.: Organic Chemistry in Colour, pp. 6–8. Springer, Berlin (1983)
4. Abrahart, E.N.: Dyes and Their Intermediates, p. 1. Chemical Publishing, New York (1977)
5. Travis, A.S.: Perkin's Mauve: Ancestor of the organic chemical industry. Technol. Cult. **31**(1), 51–82 (1990)
6. Meth-Cohn, O., Smith, M.: What did W. H. Perkin actually make when he oxidised aniline to obtain Mauveine? J. Chem. Soc. Perkin Trans. **1**, 5–7 (1994)
7. Cova, T., Pais, A., Melo, J.S.: Reconstructing the historical synthesis of Mauveine from Perkin and Caro: Procedure and details. Sci. Rep. **7** (2017). <https://doi.org/10.1038/s41598-017-07239-z>

Maxwell, James Clerk

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA



Maxwell, James Clerk

Biography

James Clerk Maxwell was born on June 13, 1831, in Edinburgh, Scotland, to a family of comfortable means and was the only child of his parents. He is considered to be a pioneer in several fields of science. Maxwell contributed greatly to the field of optics and the study of color vision and helped lay the foundations for practical color photography. Over a period of 17 years from 1855 to 1872, he published a series of papers concerning the perception of color, color blindness, and color theory [1]. He died in Cambridge at the age of 48 of abdominal cancer in 1879 [2].

Maxwell had a keen intellect and an unquenchable curiosity from childhood. Maxwell's formal schooling began unsuccessfully, and it is reported that he was treated harshly by his private tutor for being slow and disobedient [3]. He was then sent to the prestigious Edinburgh Academy to continue his education. At the age of 13, he won the school's mathematical medal as well as the first prize for English and poetry. At the age of 14, he wrote his first academic paper on *Oval Curves* which was presented, on his behalf, at the Royal Society of Edinburgh.

When 16 years old Maxwell was enrolled at the University of Edinburgh. There, among many other things, he learned about disk color mixture (Fig. 1) from one of his professors, James D. Forbes (1809–1868), who was working on a color classification system. During his undergraduate studies Maxwell also examined the properties of polarized light [4]. At age 18, he contributed two papers to the Transactions of the Royal Society of Edinburgh. One of these, *On the Equilibrium of Elastic Solids*, laid the foundation for an important discovery on the temporary double refraction produced in viscous liquids by shear stress [5]. In October 1950, Maxwell moved to the University of Cambridge's Trinity College where he graduated in 1854 with a degree in mathematics. He stayed at the Trinity College after graduation until 1856 when he accepted the position of professor of natural philosophy at Marischal College in Aberdeen [6]. In 1858 he married Katherine Mary Dewar who regularly assisted him in his experimental work. In 1860 he was granted the

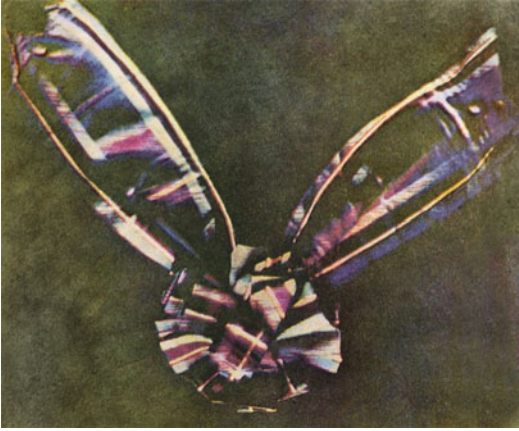


Maxwell, James Clerk, Fig. 1 Young Maxwell demonstrating one of his spinning color wheels

chair of natural philosophy at King's College in London. He stayed there until 1865, working primarily on electromagnetism. He and his wife returned to his inherited estate in Glenlair, Scotland, until he became the first Cavendish Professor of Physics at Cambridge University, where he stayed until his untimely death. Maxwell's famed equations on electro magnetism, "*Dynamical theory of the electromagnetic field*" were delivered to Michael Faraday at his residence at the Royal Society in 1864, where as a young man Faraday himself had tried on several occasions to get his work noticed. While the two men were not close, they strongly respected each other. He is considered to be the "father of electromagnetics." He is also the inventor of a thought experiment on thermodynamics, resulting in what is known as "Maxwell's demon."

Maxwell and Color

The nature of perception of color was one of Maxwell's interests which had begun in Scotland. Using an improved version of a disk mixture



Maxwell, James Clerk, Fig. 2 Reproduction of the Scottish tartan image using the photographic images and color filters employed by Maxwell and Sutton in their demonstration in 1861

apparatus, he demonstrated that white light results from a mixture of red, green, and blue light. His paper *Experiments on Colour* was a fundamental study of the color-mixing principles and was presented to the Royal Society of Edinburgh in 1855 [3]. In the following years he built a visual spectrometer in which he could mix and adjust spectral lights. The results of his own and his wife's mixture data, published in 1860, demonstrated that only three primary lights are required to match any spectral or composite light. In the same year he was awarded the Royal Society's Rumford Medal "for his researches on the composition of colours and other optical papers."

Based on his trichromatic theory, Maxwell proposed in the late 1850s a method for practical color photography. He suggested that if a scene is photographed three times separately but using red, green, and blue filters, and the resulting black and white transparent images are then superimposed on a screen using projectors equipped with similar filters, the result would be a full color reproduction of the image as seen by the human eye. Together with the photographer Thomas Sutton, he demonstrated this in 1861 at the Royal Society. As the image of the Scottish tartan shows, due to the fact that the filters were less than optimal, the outcome was less than

perfect, and commercial color photography required another 45 years of invention (Fig. 2).

References

1. Johnson, K.: *Colour Vision*. University of St. Andrews (2012). Retrieved 8 Oct 2014
2. Harman, P.M.: *Oxford Dictionary of National Biography*, vol. 37. Oxford University Press, Oxford (2004). ISBN 0-19-861411-X
3. Tolstoy, I.: *James Clerk Maxwell: A Biography*. University of Chicago Press, Chicago (1982). ISBN 0-226-80787-8
4. Mahon, B.: *The Man Who Changed Everything – The Life of James Clerk Maxwell*. Wiley, Hoboken (2003). ISBN 0-470-86171-1
5. Timoshenko, S.: *History of Strength of Materials*. Courier Dover Publications, New York (1983). ISBN 978-0-486-61187-7
6. Glazebrook, R.T.: *James Clerk Maxwell and Modern Physics*. 811951455. OCLC 811951455 (1896)

Mayer, Tobias

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Tobias Mayer, born February 17, 1723, in Marbach, Germany, a mathematician, astronomer, cartographer, and physicist, was the only child of a fountain builder and his wife. Mayer was 10 years old when his father passed away, and he grew up in impoverished circumstances in the nearby city of Esslingen, spending some time in an orphanage. He taught himself mathematics and in his later teens earned some money teaching it. Mayer also showed early interest and capabilities in drawing and painting. He moved to Augsburg to obtain more training, working for the engraver and publisher J. A. Pfeffel. At 18 he wrote and published an elementary book on mathematics and at 22 a much more detailed work on the same subject [1]. His capabilities and the fact that he had designed an accurate map of Esslingen, later published, provided for him in 1746 an offer to join the firm of the cartographer Homann in Nürnberg, at the time perhaps the most important map publisher in Europe. In 1751 he married Maria Gnüge, and a year later their son Johann Tobias was born, in his later life a mathematician/physicist. In the same year Tobias Mayer's reputation as a scientist resulted in an offer of a professorship in economy and mathematics at the University of Göttingen, where he remained until his untimely death on February 20, 1762 due to a typhus infection. In 1754 he also became the supervisor of the Royal Observatory of Göttingen, built some 10 years earlier. His major scientific achievements are a data table of the moon and highly detailed drawings of the surface of the moon based on a new methodology to achieve high accuracy and the development of a new much more accurate methodology for the determination of longitude. The latter effort resulted in his widow receiving a 3,000-pound Sterling Award from the British Parliament. One of the craters of the moon is named "T Mayer" [2].

Major Accomplishments/Contributions

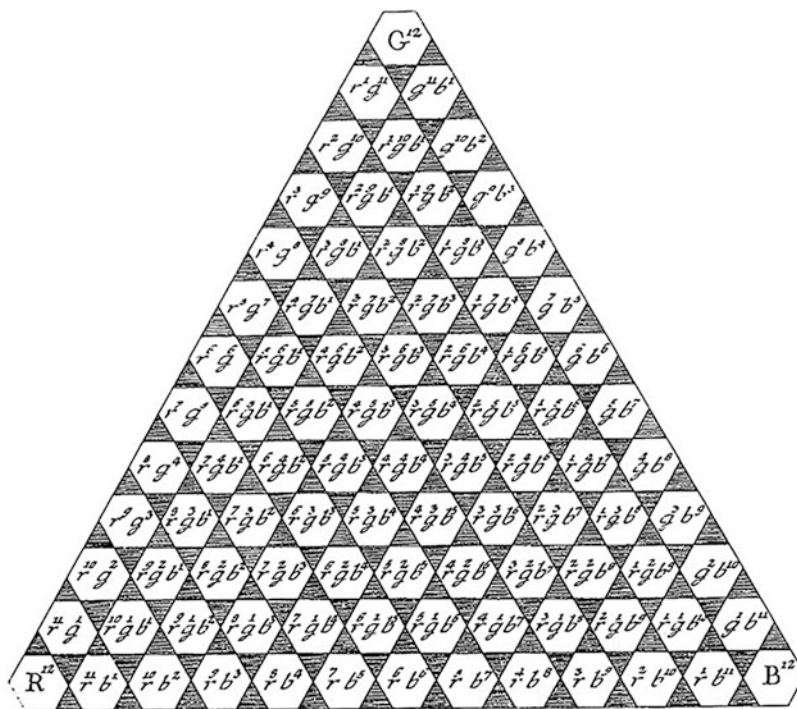
Mayer's combined interests in mathematics and painting resulted in an attempt to develop a

mathematical model of the relationships between colors. Not accepting, on practical grounds, Newton's idea of seven main colors in the spectrum, Mayer decided to base his model on the painter's primaries yellow, red, and blue. He placed these at the corners of an equilateral triangle and divided the lines between them into 12 presumably perceptually equal parts which, he assumed, could be represented by corresponding weights of the pigments used as primaries. He believed that these grades represented perceptually visible and comparable differences. The interior of the triangle is filled with colors that are mixtures of all three primaries. All of them are identified numerically, such as $r^8g^2b^2$ consisting of eight parts of the R primary and two parts each of the G and B primaries (Fig. 1) [3]. The 91 colors of the triangle of Fig. 1 can also be lightened or darkened, by addition of white or black pigments, again in twelve increments each. As these scales must end in single white or black, he gave the resulting solid the form of a double triangular pyramid, with a total of 819 defined colors [3]. In his German translation of the original Latin text, H. Lang defined Mayer's system as attempting to meet five then novel conditions: (1) All possible object colors are represented. (2) All colors are the result of mixtures of three primary colors. (3) Each color is represented by a triple number indicating the content of the three primaries. (4) The totality of all colors in the system is contained in a three-dimensional color solid. (5) The differences between neighboring colors correspond to perceptual differences [4].

Mayer presented his ideas in 1758 in a public meeting of the Society of Sciences in Göttingen, a report of which was published in the newspaper *Göttingische Anzeigen von gelehrten Sachen* (Göttingen reports on scholarly matters) some three weeks later [5]. This report was the basis of J. H. Lambert's work on his triangular color pyramid [6]. Mayer had not done much experimental work to implement his system. As pointed out by Lambert, Mayer had not been aware of the varying coloristic strength of the three primaries, requiring consideration for the purpose of obtaining perceptual uniformity.

Mayer, Tobias,

Fig. 1 Central plane of Mayer's double triangular pyramid color solid



In 1958 Mayer also invented a new coloration method for prints. He proposed, and produced an example of, the coloration to consist of sections of wax containing different amounts of pigments to achieve different colors [7]. Multiple prints could then be produced from the wax collage. The idea proved to be too complex to be practical and was not pursued after his death.

Haude und Spener, Berlin (1772). English translation available at www.iscc.org

7. Günther, G.C.: *Praktische Anweisung zur Pastellmalerei*, pp. 129–132. Schneider, Nürnberg (1792)

M

References

1. Mayer, T.: *Mathematischer Atlas*. Pfeffel, Augsburg (1745)
2. Forbes, E.G.: The life and works of Tobias Mayer. *Q. J. R. Astron. Soc.* **8**, 227–251 (1967)
3. Mayer, T.: *De affinitate colorum commentatio* (On the relationship between colors), published posthumously in G. C. Lichtenberg, *Opera inedita Tobiae Mayeri* (Unpublished works of Tobias Mayer). Dieterich, Göttingen (1775). English translation: On the relationships between colors (A. Fiorentini, transl.), *Color Research and Application* **25**, 66–73 (2000)
4. Lang, H.: Tobias Mayers Abhandlung über die Verwandtschaft der Farbe. *Die Farbe*. **28**, 1–34 (1980)
5. *Göttingische Anzeigen von gelehrten Sachen*, part 147, p. 1385–9 (1758)
6. Lambert, J. H.: Beschreibung einer mit dem Calauischen Wachse ausgemalten Farbenpyramide.

Mechanisms of Color Categorization

Yasmina Jraissati

American University of Beirut, Beirut, Lebanon
Ronin Institute, NJ, USA

Synonyms

Factors of color categorization; Origin of proto-types; Origin of boundaries

Definition

Color categories vary across cultures, but there are important regularities. Where do these come from? Color categories are graded and feature a

focal color or prototype. In the literature, for a long time, the question pertaining to the origin of color categories was conflated with the role and nature of prototypes. This entry goes over the main answers to this question.

What Is Color Categorization and Why It Matters

Some have claimed that categories are the building blocks of cognition [1]. Our experience of the world is temporal, featuring a series of events and sensory experiences. Yet, we are able to sort and organize these sensations into groups called categories. In the case of color, we categorize color experiences on the basis of the similarity of their chromatic qualities. The way colors are categorized is not trivial. Apart from their use in memory, action, and communication, categories are also useful in inference and induction.

While I have a concept of red, a mental representation that encompasses contents like “red strawberries are sweet,” the category RED is not a mental representation and consists of a group of items in the world, to which the label “red” applies. However, things that are red at different degrees are all equally called “red.” How then, does an agent categorize? More generally, the question pertains to how a given categorization system emerges ([2] for a recent review).

Graded Membership and Prototype Theory

According to Rosch, there are two cognitive principles of categorization: cognitive economy and perceived world structure. The generally accepted hypothesis, which continues to inform recent categorization models, is that a category system provides maximum information with the least cognitive effort: There is a trade-off between having a large number of categories with as fine discriminations between categories as possible and having a small number of categories that will reduce cognitive effort. The cognitive economy principle tends toward reducing the infinite

differences among objects to proportions that are cognitively and behaviorally usable. On the other hand, the perceived world structure principle tends to gather information from the world, granting that the world possesses high correlational structure – e.g., wings tend to co-occur with feathers more than with fur.

Given perceived world structure, it follows that certain cues predict category membership: A given cue x is a predictor of a category y (e.g., feathers are a predictor of birds). As the frequency with which cue x is associated with category y increases, the validity of x as a predictor of y increases. The validity of x decreases as the frequency with which x is associated with categories other than y increases (e.g., fur is *not* a good predictor of birds). Consequently, the entire category has as a cue validity: It is the sum of the cue validities of all the attributes of the category. Thus, by definition, a category with high cue validity will be more differentiated from other categories than a category of lower cue validity.

There are two dimensions along which objects can be categorized: The “vertical” dimension refers to the inclusiveness of a category, while the “horizontal” dimension refers to the segmentation of the world at the same level of inclusiveness. So, for example, bird and robin are on the same vertical dimension (a robin is a kind of bird), while cat and dog are on the same horizontal dimension. On the vertical dimension, members of superordinate categories (e.g., mammals) have lower total cue validity, as the number of shared attributes is low. Members of subordinate categories also have lower total cue validity (e.g., Dalmatians), as the number of attributes shared with other categories (e.g., Beagles) is high.

The cue validity is maximized at the so-called *basic* level (e.g., dog). Similarly, on the horizontal dimension, within a category, some members of the category have higher cue validity than others and are better predictors of the category. These members of the category share more attributes with more members of the category and are, therefore, more similar to a larger number of category members than others. These are most representative, *prototypical* members of a category.

Basicness and prototypicality result, according to Rosch, from the two cognitive mechanisms of cognitive economy and perceived world structure. Prototype theory describes a categorization process and the resulting category structure. It also accounts for the cognitive advantages that prototypes have in comparison to other category members (e.g., typicality effects: prototypes are processed faster, learned faster, and produced faster). Prototype theory does *not* however offer an ontology: Prototypes should not be reified or identified with a specific category member or mental structure. People's judgments of goodness of membership is what defines prototypes. Thus, prototypes, and also categories, are context dependent.

Color Categories

In 1969 Berlin and Kay published a seminal monograph, *Basic Color Terms: Their Universality and Evolution* [3], in response to Boas [4] and Whorf [5], who had argued that notwithstanding the universality of human perceptual apparatus, color lexical categories varied across languages. In their perspective, language is the reason why we categorize the way we do [6, 7].

Led by the simple observation that some color terms in English translated well in many languages, Berlin and Kay conducted a cross-cultural naming task. They compared a total of 98 languages and made the following claims: (1) There is a set of 11 universal basic color categories; (2) color terms emerge in lexicons following a constrained evolutionary sequence.

These claims are somewhat controversial because what counts as a "basic" color term may be intuitive but is difficult to define in a way that is not ad hoc [8–10]. Also, different languages may have different numbers of color terms, but this is not enough to talk of lexical "evolution." The remarkable observation in *Basic Color Terms* (BCT) and the subsequent World Color Survey (WCS) is that unrelated languages, featuring the same number of basic color terms, have categories that partition color space in similar ways. Thus, a language featuring only four basic color terms and

categories, usually features color terms referring to the category BLACK (extending over dark colors), WHITE (light colors), WARM (red, yellow and orange), and COOL (blue and green). A language featuring six basic color terms will typically have terms referring to BLACK, WHITE, RED, YELLOW, BLUE, and GREEN.

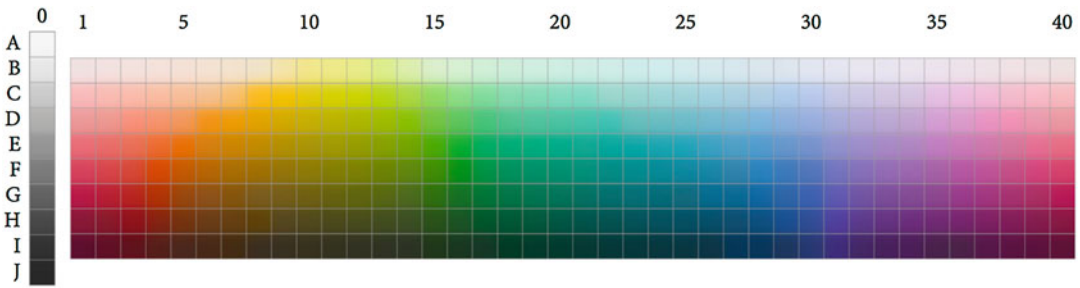
Where Do Prototypes Come from? The Perceptual Account

As in the case of birds or dogs, color categories feature a graded structure. Take the category we call "blue" in English. Some colors are focal or most typical of "blue," while others could be labeled "green." In a two-dimensional representation of color (Fig. 1), categories are represented as an irregularly ovoid area (Fig. 2). Focal colors figure toward the center of the category: This is where membership is highest (Fig. 3), and where typicality effects can be observed [11, 12]. Membership gradually decreases and is lowest at the periphery [13].

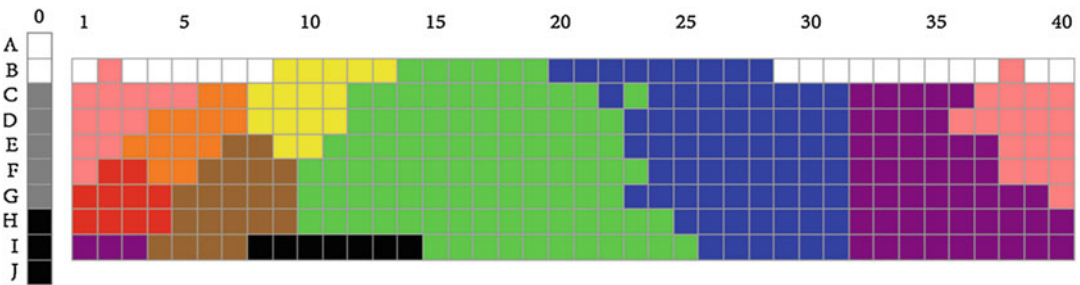
What is the role of prototypes and where do they come from? Because of the sensory nature of color categories, and encouraged by the color vision science of the 1970s [14], Rosch hypothesized that color categories were shaped by naturally determined prototypes [15]. Thus, in the case of color, Rosch supported the reification of prototypes. It was believed that the output of the second stage opponent color processing mechanism yielded none other but Hering's unique hues – white, black, red, green, yellow, and blue [16]. Interestingly, WHITE, BLACK, RED, YELLOW, GREEN, and BLUE are the first six categories of the BCT's evolutionary sequence. The suggestion was therefore that basic color categories' prototypes resulted from these neuronal responses [17].

Ecological Constraints

A different account of prototypes is sought in ecology. Based on images of natural scenes, amounting to 10,000 pixels, a process of



Mechanisms of Color Categorization, Fig. 1 The Munsell array used in the WCS



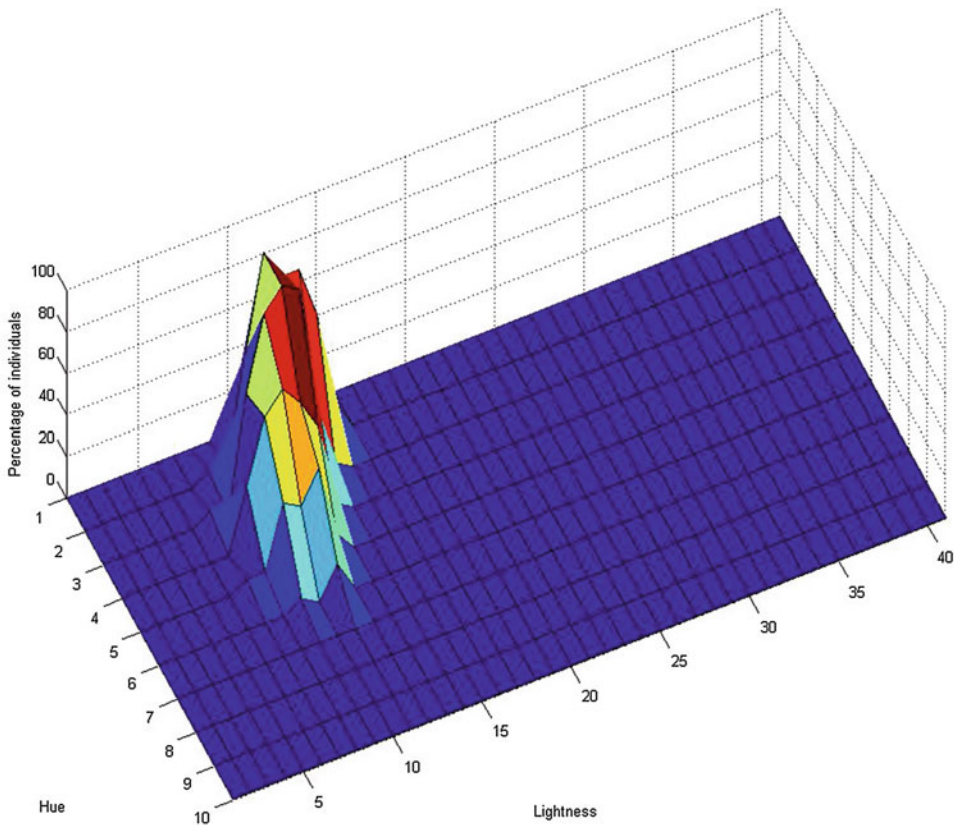
Mechanisms of Color Categorization, Fig. 2 Mode map of the French lexicon [42]

categorization was modeled through grouping natural image statistics by a clustering algorithm, with grouping based on the minimum-distance criterion [18]. It appears that the centers of these clusters are most similar to the configuration of natural languages' focal colors. The distribution of colors in the world, as they were experienced in the past would in this approach account for the location of focal colors in color space. Also, a high correlation was found between the order of emergence of cluster centers and the constrained evolutionary sequence offered by the BCT. A trade-off between accuracy in representation and simplicity of the category system is thought to guide the number of categories in a given lexicon (also see [19]).

However the role of ecology in categorization is a complex question, and there is more than one possible direction of influence. It is possible that color distribution in the environment shapes color categorization, but the structure of our perceptual experience might itself result from an adaptation

to ecology. For example, Shepard [20] contended that the reason why light/dark, red/green, and yellow/blue are the dimensions of our perceptual experience has to do with the variation of the color of light. This three-dimensional representation of color may result from an adaptation of our light processing system [21] to a pervasive and enduring feature of the world (also see [22]). This three-dimensional nature of color experiences would in turn translate into categorical perception. Indeed, there is evidence of a form of categorical perception in animals [23]; and when it comes to pre-linguistic human infants, there is also mounting evidence of categorical behavior [24–26].

However, the implications of these observations when it comes to factors underlying human adult categorization are not straightforward. Importantly, English infant categories do not correspond to those of adult English speakers [27]: Unlike adult categories, their boundaries coincide with the cone-opponent dimensions resulting from the second-stage mechanism, which are in



Mechanisms of Color Categorization, Fig. 3 Heat map of membership to “jaune” as measured by consensus in naming. (From [42], unpublished)

fact closer to cherry/chartreuse and orange/teal than to yellow/blue and red/green [26, 28].

Perceptual Mechanisms Are Not Enough

Though it provides a powerful account of color categorization, a biological approach to prototypes faces many challenges. First, we know today, as just mentioned, that the outputs of the second-stage opponent process mechanism does not yield the unique hues at all [28]. Second, reducing prototypes to neuronal responses might account for their universality, but it does not explain lexicon evolution. Third, if prototypes are naturally given, it is surprising why not all languages have categories for each unique hue.

Finally, the biological reduction of color prototypes is conceptually unappealing: If color categories were biologically fixed, there would be no trade-off between maximum information and minimum effort in a given context.

These are the main reasons why subsequent models of color categorization break away from this kind of strong biological approach. In a seminal paper, Regier, Kay, and Khetarpal [29] explicitly aimed at a model that would make room for cross-linguistic variations. Building on a proposal by Jameson and D’Andrade [30], they make use of the idea of the irregularity of perceptual space: Our perceptual sensitivity thresholds are not uniform across the color spectrum resulting into some areas being perceptually more salient than others. Given an endowed capacity to produce

“optimal” categories (members of a category are most similar to each other, and least similar to members of other categories), the irregularity of perceptual space would lead to universal patterns of categorization for various numbers of categories. This model is promising as it is verified for up to six categories ($n = 1, n = 2 \dots, n = 6$), but it does not predict the categorization pattern of $n = 11$ [31], knowing that many languages feature 11 basic color categories or more (e.g., English, French, Turkish, Arabic, etc.).

Connectedly, a series of recent studies has shown that the relationship between perception and lexical categorization was only loose [32, 33]. Testing just noticeable difference thresholds across an isoluminant hue circle in DKL space (which represents sensitivities of the opponent mechanisms), these studies have shown that only two lexical boundaries coincide exactly with sharp dips in sensitivity thresholds, namely, the lexical boundaries “blue”/“green” and “pink”/“yellow” (there was no category for “red” at the selected brightness level). Together, these results suggest that there must be more to lexical categorization than perceptual mechanisms, thereby re-introducing linguistic and cultural factors.

Communication Constraints and Cultural Transmission

It should be noted that linguistic relativism does not deny perceptual constraints: A color category cannot be discontinuous in color space, it can only include neighboring colors in its extension. Also, the same colors are considered as neighboring across cultures: This assumes that humans experience the same continuum, which ultimately results from the overlapping sensitivity ranges of their photoreceptors [34]. The main point of disagreement culminates in the nature and role of prototypes. The contention here is that prototypes result from, and do not precede, category formation, as in Rosch’s initial and general account. That is, in language use, a consensual extension of a given color term stabilizes, and consequently, the prototypical area emerges as an abstraction of a central tendency.

Prototypes should however not be understood merely as centroids that result from the category’s extension: Indeed, category boundaries vary more than focal colors do [35]. Instead, some have suggested that the best examples are most representative, in the sense where prototypes reflect both high similarity to other members of the category and low similarity to members of other categories [36]. This would account both for regular patterns of categorization and for unusual ones. What determines color categories in this view is the functional pressure for informative communication given the structure of color space [37]. From this information-theoretic perspective, efficiency in communication leads to need (reflecting the world’s structure) and precision (resulting from perceptual structure) being connected. Ultimately, this approach suggests that language in communication and perceptual structure both shape color categorization [38].

Thus, it should not be forgotten that color lexica are ultimately a population phenomenon. A seminal study [39] using hundreds of artificial agents and Munsell colors shows that linguistic communication and other non-linguistic social interactions coordinate color categorization, even in the absence of a statistical structure to the data, that is, regardless of the distribution of color in the environment. Other population factors, like birth-death cycles [40], play a role in reaching a consensual and stable categorization system, similar to the ones observed in natural languages. A link between the lifespan of the population and the number of color terms is suggested. Interestingly, birth-death cycles in human societies vary with technological development, and, incidentally, so does the size of color lexicons. The important input of these studies is that beyond communication, cultural transmission might more specifically be at play [41].

Conclusion

Color categories are graded: Some objects have a higher membership degree than others. According

to Rosch, best members of the category are the prototypes, which are taken to result from two principles: Cognitive economy and perceived world structure. In the case of color however, the nature of the prototypes has been debated. The way they are determined is a central question, as it critically connects to the more general understanding of the origin of color categories.

Cross-References

- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Conceptual Spaces](#)
- [Cross-cultural Communication](#)
- [Hering, Karl Ewald Konstantin](#)
- [Unique Hues](#)

References

1. Harnad, S.: Chapter 2 – To cognize is to categorize: cognition is categorization. In: Cohen, H., Lefebvre, C. (eds.) *Handbook of Categorization in Cognitive Science*, 2nd edn, pp. 21–54. Elsevier, San Diego (2017)
2. Lindsey, D.T., Brown, A.M.: Recent progress in understanding the origins of color universals in language. *Curr. Opin. Behav. Sci.* **30**, 122–129 (2019). <https://doi.org/10.1016/j.cobeha.2019.05.007>
3. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. CLSI Publications, Berkeley (1969)
4. Boas, F.: *The Mind of Primitive Man*. The Macmillan Company, New York (1938)
5. Whorf, B.-L.: Science and linguistics. *Technol. Rev.* **42**, 229–231, 247–248 (1940)
6. Davidoff, J., Goldstein, J., Roberson, D.: Nature versus nurture: the simple contrast. *J. Exp. Child Psychol.* **102**, 246–250 (2009)
7. Evans, N., Levinson, S.C.: The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448 (2009). <https://doi.org/10.1017/S0140525X0999094X>
8. Crawford, T.D.: Defining “basic color term”. *Anthropol. Linguist.* **24**, 338–343 (1982)
9. Hickerson, N.: Review of basic color terms : their universality and evolution. *Int. J. Amercian Linguist.* **37**, 257–270 (1971)
10. Jraissati, Y.: Proving universalism wrong does not prove relativism right: considerations on the ongoing color categorization debate. *Philos. Psychol.* **27**, 401–424 (2014)
11. Boynton, R.M., Olson, C.X.: Locating basic colors in the OSA space. *Color. Res. Appl.* **12**, 94–105 (1987)
12. Guest, S., van Laar, D.: The structure of colour naming space. *Vis. Res.* **40**, 723–734 (2000)
13. Hampton, J.A.: Typicality, graded membership, and vagueness. *Cogn. Sci.* **31**, 355–384 (2007)
14. Jameson, D., Hurvich, L.M.: Some quantitative aspects of an opponent-colors theory I. Chromatic responses and spectral saturation. *J. Opt. Soc. Am.* **4**, 546–552 (1955)
15. Rosch, E.: Natural categories. *Cogn. Psychol.* **4**, 328–350 (1973)
16. De Valois, R., Abramov, I., Jacobs, G.H.: Analysis of response patterns of LGN cells. *J. Opt. Soc. Am.* **56**, 966–977 (1966)
17. Kay, P., McDaniel, C.K.: The linguistic significance of the meanings of basic color terms. *Language.* **54**, 610–646 (1978)
18. Yendrikhovskij, S.N.: A computational model of colour categorization. *Color. Res. Appl.* **26**, 235–238 (2001)
19. Yendrikhovskij, S.N.: Computing color categories from statistics of natural images. *J. Imaging Sci. Technol.* **45**, 409–417 (2001)
20. Shepard, R.N.: Perceptual-cognitive universals as reflections of the world. *Psychon. Bull. Rev.* **1**, 2–28 (1994)
21. Osorio, D., Vorobyev, M.: A review of the evolution of animal colour vision and visual communication signals. *Vis. Res.* **48**, 2042–2051 (2008). <https://doi.org/10.1016/j.visres.2008.06.018>
22. Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S.T., Conway, B.R.: Color naming across languages reflects color use. *Proc. Natl. Acad. Sci.* **114**, 10785–10790 (2017). <https://doi.org/10.1073/pnas.1619666114>
23. Siuda-Krzywicka, K., Boros, M., Bartolomeo, P., Witzel, C.: The biological bases of colour categorisation: from goldfish to the human brain. *Cortex.* **118**, 82–106 (2019). <https://doi.org/10.1016/j.cortex.2019.04.010>
24. Franklin, A., Davies, I.R.L.: New evidence for infant color categories. *Br. J. Dev. Psychol.* **22**, 349–377 (2004)
25. Clifford, A., Franklin, A., Davies, I.R., Holmes, A.: Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain Cogn.* **71**, 165–172 (2009)
26. Skelton, A.E., Catchpole, G., Abbott, J.T., Bosten, J. M., Franklin, A.: Biological origins of color categorization. *Proc. Natl. Acad. Sci.* **114**, 5545–5550 (2017). <https://doi.org/10.1073/pnas.1612881114>
27. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R. L., Kay, P., Regier, T.: Categorical perception of color

- is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *PNAS*. **105**, 3221–3225 (2008)
28. Abramov, I., Gordon, J.: Color appearance: on seeing red – or yellow, or green, or blue. *Annu. Rev. Psychol.* **45**, 451–485 (1994)
 29. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. *PNAS*. **104**, 1436–1441 (2007)
 30. Building on a proposal by Jameson, K., D’Andrade, R. G.: It’s not really red, green, yellow, blue: an inquiry into perceptual color space. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 295–320. Cambridge University Press, Cambridge (1997)
 31. Jraissati, Y., Douven, I.: Does optimal partitioning of color space account for universal color categorization? *PLoS One*. **12**, e0178083 (2017)
 32. Witzel, C., Gegenfurtner, K.R.: Categorical sensitivity to color differences. *J. Vis.* **13**, 1–33 (2013)
 33. Witzel, C., Gegenfurtner, K.R.: Are red, yellow, green, and blue perceptual categories? *Vis. Res.* **151**, 152–163 (2018)
 34. Roberson, D., Davies, I., Davidoff, J.: Color categories are not universal : replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
 35. Regier, T., Kay, P., Cook, R.: Universal foci and varying boundaries in linguistic color categories. In: *The 27th Reunion of the Cognitive Science Society*, pp. 1827–1832, Stressa (2005)
 36. Rosch, E., Mervis, C.B.: Family resemblances: studies in the internal structure of categories. *Cogn. Psychol.* **7**, 573–605 (1975)
 37. Abbott, J.T., Griffiths, T.L., Regier, T.: Focal colors across languages are representative members of color categories. *Proc. Natl. Acad. Sci.* **113**, 11178–11183 (2016)
 38. Zaslavsky, N., Kemp, C., Regier, T., Tishby, N.: Efficient compression in color naming and its evolution. *Proc. Natl. Acad. Sci.* **115**, 7937–7942 (2018). <https://doi.org/10.1073/pnas.1800521115>
 39. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language. A case study for colour. *Behav. Brain Sci.* **24**, 469–529 (2005)
 40. Park, J., Tauber, S., Jameson, K.A., Narens, L.: The evolution of shared concepts in changing populations. *Rev. Philos. Psychol.* (2018). <https://doi.org/10.1007/s13164-018-0420-5>
 41. Silvey, C., Kirby, S., Smith, K.: Communication increases category structure and alignment only when combined with cultural transmission. *J. Mem. Lang.* **109**, 104051 (2019). <https://doi.org/10.1016/j.jml.2019.104051>
 42. Claidière, N., Jraissati, Y., Chevallier, C.: A colour sorting task reveals the limits of the universalist/relativist dichotomy: colour categories can be both language specific and perceptual. *J. Cogn. Cult.* **8**(3–4), 211–233 (2008)

Melanin

Tova N. Williams

Wilson College of Textiles, North Carolina State University, Raleigh, NC, USA

Synonyms

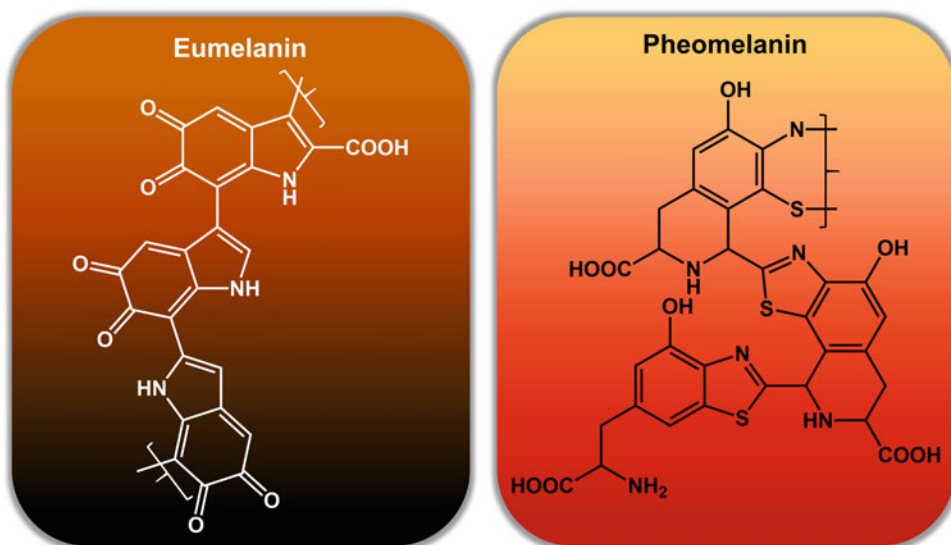
Natural Hair Colorant; Natural Hair Pigmentation

Definition

Melanin is the natural colorant in protein substrates such as hair, but it makes up only a small percentage of the hair fiber’s mass. Two types exist – eumelanin and pheomelanin – which, depending on factors such as concentration, can lead to colors including black, brown, red, and blonde.

Overview of Melanin Pigment

Melanin is a colorant found in biological systems including plants, animals, molds, and mildews, as well as the human body [1]. It is a pigment that is a component of the human eye, skin, and hair. In hair, it is found almost exclusively in the cortex (or the inner region) [1, 2]. Although melanin pigment constitutes less than 3% of human hair fiber mass [3], it plays prime importance in defining an individual. Natural hair color is mandated by an individual’s parents, and thus, is a reflection of one’s ancestry [4]. Natural hair color will change over time – accidentally or naturally. For example, a person experiencing a severe iron deficiency may notice a shift in black to brown color of his or her hair [3]. And as a part of aging, hair will produce less pigment and become gray [5, 6], coarse, and resistant to dyeing [3]. Completely unpigmented hair is white [7], and the scattering of light from the exterior of hair lacking pigment is responsible for its gray/white color [6]. Presently, the mechanisms for hair depigmentation are



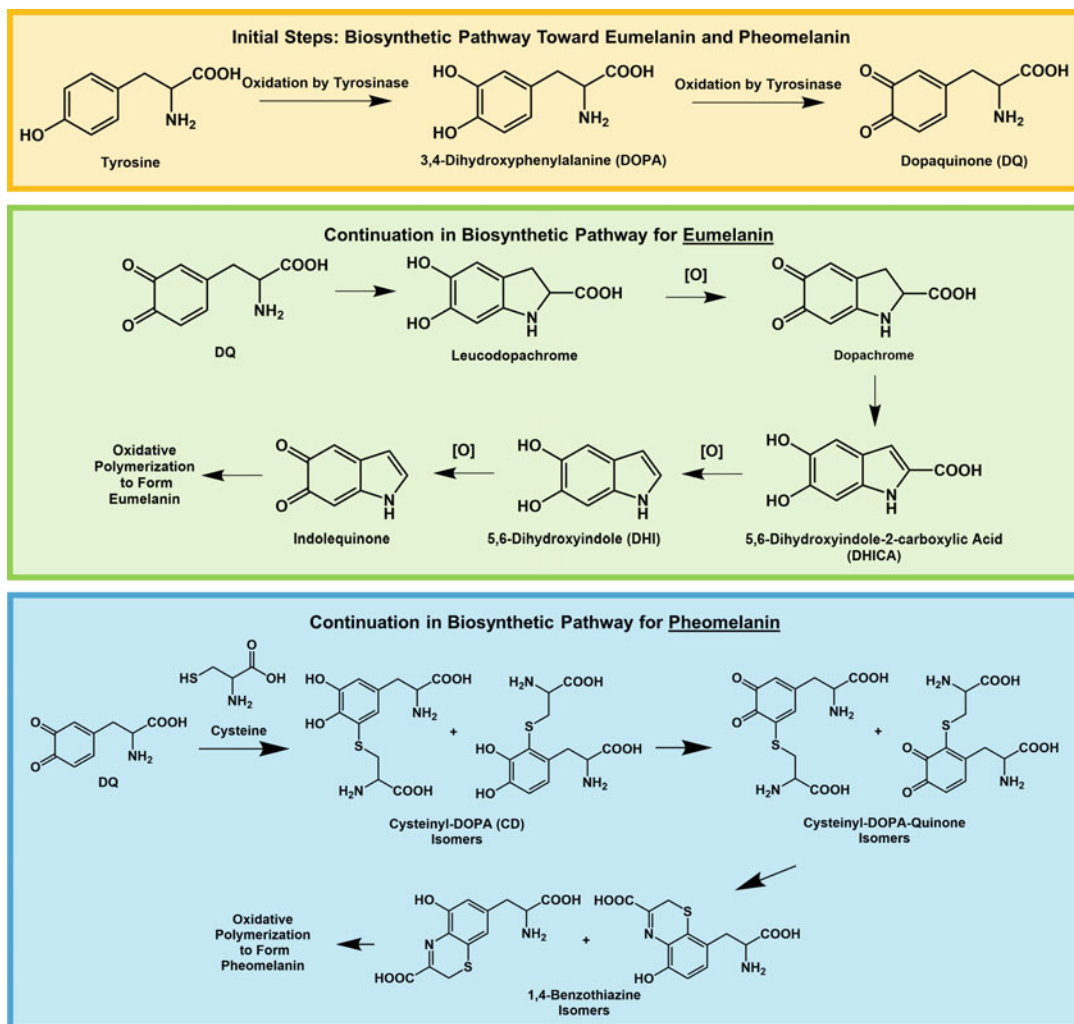
Melanin, Fig. 1 Proposed structures of eumelanin and pheomelanin pigments as determined by Gray [3]

being investigated, but it seems the mostly likely causes are hereditary factors.

Melanocytes are cells that secrete melanosomes during hair growth and development [3, 6]. The melanosomes initially contain protein matter (melanoprotein) that can bind melanin pigment [8, 9]. The pigment will eventually be transferred to other cells [3], for example, keratinocytes (or keratin-producing cells) [6, 8, 9]. Fascinatingly, nature only needs two substances – eumelanin and pheomelanin – to produce a variety of shades from black to brown to red to blonde unlike the hair color chemist who requires far more colorants [8]. Eumelanin and pheomelanin (see Fig. 1 for proposed structures) are contained within melanosomes and estimated to be 0.2–0.8 μm in length [3]. They can form aggregates of about 300 nm [10]. Eumelanin gives brown to black shades [1, 3, 6], and the melanosomes housing it have been described as ellipsoidal in shape [3]. Contrastingly, others have described the shape of the melanosomes as ovoidal or spherical with uniform shape and defined edges [8, 9]. Pheomelanin, sometimes referred to as “phaeomelanin” [3], gives red to blonde shades [1, 3, 6, 9]. The melanosomes housing it have been described as spherical [3]. Others have described these melanosomes as moderately

spherical or rod-like [8, 9] and smaller than those of eumelanin [8]. The most dominant pigment seen in the global population is eumelanin [3, 6], found most exclusively in Japanese [8, 9, 11] and Korean hair [11]. On the other hand, pheomelanin is the dominant pigment found in Celtic hair [8, 9, 11]. Black hair contains a higher concentration of pigment deposited in melanosomes, and this pigment is more uniformly dispersed [3]. In fact, the darker the hair, the more concentrated the eumelanin [8, 9]. Melanin particle size and granule density contribute to the depth of shade achieved, while the proportion of the melanin types contributes to the shade itself [12]. Over the lifespan of an individual, the type and concentration of melanin will change [3]. Since not all hair follicles produce the same types and concentration of melanin, variation is even evident spanning the different regions of the body.

Melanins are believed to be amorphous and polymeric. However, their absolute structures have been difficult to determine due to their insolubility [1]. The formation of melanins takes advantage of bifunctional enzymes that catalyze the hydroxylation of phenols and subsequent oxidation of diols to quinones. In the present case, the biosynthetic routes for



Melanin, Fig. 2 Biosynthetic pathways toward eumelanin and pheomelanin [8, 13]

eumelanin and pheomelanin have been proposed to be similar, beginning with a pair of tyrosinase enzyme-induced oxidation reactions that involve the conversion of tyrosine to 3,4-dihydroxyphenylalanine (DOPA) and then to dopaquinone (DQ) [1, 3, 8, 9] (see Fig. 2). Inside the melanosomes, DQ undergoes cyclization to leucodopachrome followed by oxidation to the orange-colored compound, dopachrome. Dopachrome undergoes isomerization to form the carboxylated diol, 5,6-dihydroxyindole-2-carboxylic acid (DHICA), followed by decarboxylation to give

the corresponding indole, 5,6-dihydroxyindole (DHI). A series of oxidation steps produces indolequinone and culminates in oxidative polymerization to give eumelanin [3]. This chemistry has been supported by chemical degradation studies designed to establish DHICA and DHI as key building blocks of eumelanins. For pheomelanin formation, the two-step conversion of tyrosine to DQ is followed by a reaction with cysteine to form a mixture of cysteinyl-DOPA (CD) isomers and then a CD-quinone mixture, which is then converted to benzothiazine and benzothiazol isomers. These

heteroaromatic products undergo polymerization to produce pheomelanin [13].

High-field (600–750 MHz) 2-D (^{13}C and ^{15}N) solid-state NMR spectroscopy has been used to characterize synthetic melanin prepared from ^{13}C and ^{15}N enriched L-tyrosine or dopamine [14]. The goal of this investigation was to label the resulting eumelanins and demonstrate that cell-free polymerization of the two precursors produces melanins with an indole-based aromatic core structure found in natural melanin pigments. While researchers had assumed that synthetic melanins were suitable models of melanin from natural sources, their structural agreement had not been demonstrated directly. Further, the use of solid-state NMR for the analysis helped overcome the investigative challenges posed by these insoluble, heterogeneous pigments.

Cross-References

► [Colorant, Natural](#)

References

1. Brown, K.C., Marlowe, E., Prota, G., Wenke, G.: A novel natural-based hair coloring process. *J. Soc. Cosmet. Chem.* **48**, 133–140 (1997)
2. Robbins, C.R.: Morphological, macromolecular structure and hair growth. In: *Chemical and Physical Behavior of Human Hair*, pp. 1–104. Springer, Berlin/Heidelberg (2012)
3. Gray, J.: Hair pigments and the hair follicle. In: *The World of Hair Colour*, pp. 19–35. Thomas Learning, Bedford Row, London (2005)
4. Gray, J.: Colour in the world. In: *The World of Hair Colour*, pp. 3–6. Thomas Learning, Bedford Row, London (2005)
5. Robbins, C.R.: Dyeing human hair. In: *Chemical and Physical Behavior of Human Hair*, pp. 445–488. Springer, Berlin/Heidelberg (2012)
6. Rust, R.C., Schlatter, H.: Hair dyes. In: Draelos, Z. (ed.) *Cosmetic Dermatology: Products and Procedures*, pp. 239–249. Wiley, Chichester (2016)
7. Popescu, C., Höcker, H.: Hair – the most sophisticated biological composite material. *Chem. Soc. Rev.* **36**, 1282–1291 (2007)
8. Morel, O.J.X., Christie, R.M.: Current trends in the chemistry of permanent hair dyeing. *Chem. Rev.* **111**, 2537–2561 (2011)
9. Christie, R.M., Morel, O.J.X.: The coloration of human hair. In: Lewis, D.M., Rippon, J.A. (eds.) *The Coloration of Wool and Other Keratin Fibers*, pp. 357–391. Wiley, United Kingdom, Bradford (2013)
10. Sampaio, S., Maia, F., Gomes, J.: Diffusion of coloured silica nanoparticles into human hair. *Color. Technol.* **127**, 55–61 (2011)
11. Im, K., Kim, T., Jeon, J.: Metal-chelation-assisted deposition of polydopamine on human hair: a ready-to-use eumelanin-based hair dyeing methodology. *ACS Biomater. Sci. Eng.* **3**, 628–636 (2017)
12. Robbins, C.R.: Genetic control/involvement in hair Fiber traits. In: *Chemical and Physical Behavior of Human Hair*, pp. 177–204. Springer, Berlin/Heidelberg (2012)
13. Ito, S., Wakamatsu, K.: Chemistry of melanins. In: Nordlund, J.J., Boissy, R.E., Hearing, V.J., King, R. A., Oetting, W.S., Ortonne, J.-P. (eds.) *The Pigmentary System: Physiology and Pathophysiology*, pp. 282–310. Blackwell, Malden (2006)
14. Chatterjee, S., Prados-Rosales, R., Tan, S., Itin, B., Casadevall, A., Stark, R.E.: Demonstration of a common indole-based aromatic core in natural and synthetic eumelanins by solid-state NMR. *Org. Biomol. Chem.* **12**(12), 6730–6736 (2014)

Melanopic Equivalent Daylight Illuminance

Wout van Bommel

van Bommel Lighting Consultant, Nuenen, The Netherlands

Definition

Illuminance produced by radiation conforming to standard daylight (D65) that provides an equal melanopic irradiance as the light source considered. Melanopic irradiance is the result of the human pRGC response due to its photopigment melanopsin. pRGCs (intrinsically photosensitive retinal ganglion cells) are those ganglion cells located in the retina of the human eye that are photosensitive by means of the photopigment melanopsin.

Photosensitive Retinal Ganglion Cells (pRGCs)

In 2002, Dave Berson and coworkers from the Brown University in the USA discovered a third type of photoreceptor in addition to the well-known cone and rod types [1]. A small part of the already long known ganglion cells in the retina are sensitive for light incident on them. This third type of light-sensitive cell is called “photosensitive retinal ganglion cell”: pRGC. In literature it is also referred to as ipRGC which stands for intrinsically photosensitive retinal ganglion cell. This particular type of ganglion cell transmits signals to the supra-chiasmatic nucleus (SCN), a pacemaker organ in the brain that provides time cues to the body. The SCN, often referred to as the biological clock, controls many body properties with a circadian rhythm (24-hour rhythm). Examples are the heart rate, blood pressure, liver function, generation of new cells, and the production of many hormones. The hormones melatonin and cortisol are of particular importance because they govern sleep and alertness. The free-running rhythm of the SCN varies between individuals from approximately 23.5 to 25 hours. The 24-hour rhythm of natural daylight-darkness synchronizes the biological clock with the 24-hour rhythm: The SCN is no longer free running. This time-giver (zeitgeber) signal reaches the SCN through the photosensitive retinal ganglion cells. A correct (24-hours) circadian rhythm is of great importance for human health, well-being, alertness, and performance.

Spectral Sensitivity of Photopigments

Melanopsin is the photopigment that gives the pRGC cell its intrinsic photosensitivity [2–5]. It is an opsin type of photopigment different from the rod opsin as well as different from the three types of cone opsins. All these different opsin types of pigments absorb different wavelengths. They, therefore, result in different spectral sensitivities for the different kinds of photocells.

The absorption spectra of the cones and rods are already long time known. Today, also the absorption spectrum of melanopsin has become available [6, 7]. CIE defines the spectral melanopic sensitivity based on measurements of the absorption spectrum of the photopigment melanopsin in its international standard 026:2018 [7]. It determines the intrinsic sensitivity of the pRGC cells without the interaction effect of rods and cones through neural wiring. Figure 1 shows the melanopic spectral sensitivity curve of CIE standard 026:2018 (full drawn, light blue curve). The maximum melanopic sensitivity is reached at a wavelength of approximately 480 nm. Figure 1 also shows the sensitivity curves for the S, M, and L cones and rods, based on their specific photopigments (dashed curves).

With these curves, it is possible to calculate for a specific lamp spectrum and a given amount of light on the outer surface of the eye, the spectrally weighted irradiance for the five human photopigments. Irradiance is the power of radiation received by a surface divided by the area of that surface. It is expressed in Watt/m^2 . The symbol for irradiance, E_e , is written with the subscript “e” to avoid confusion with the symbol E used for the photometric unit illuminance (i.e., $V(\lambda)$ weighted irradiance expressed in lux).

The spectrally weighted irradiance, for the example of melanopsin, is obtained from the formula:

$$E_{e, \text{ mel}} = \int E_e(\lambda) S_{\text{mel}}(\lambda) d(\lambda)$$

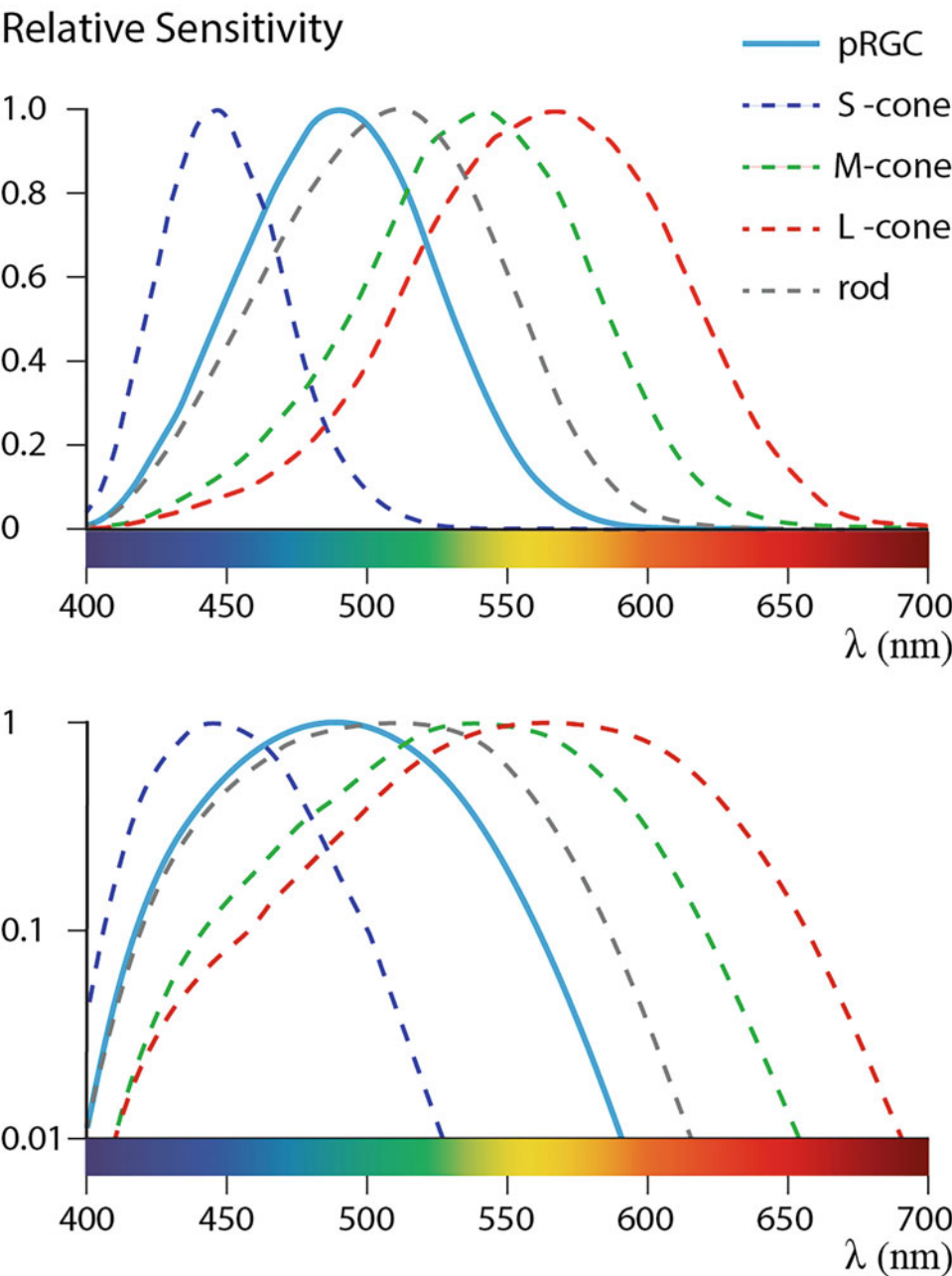
where

$E_{e, \text{ mel}}$ = melanopsin weighted irradiance in W/m^2

$E_e(\lambda)$ = irradiance for wavelength λ in W/m^2 (available from the lamp spectrum)

$S_{\text{mel}}(\lambda)$ = relative melanopsin sensitivity for wavelength λ , where $S_{\text{mel}}(\lambda)$ is normalized to the value of 1 at its peak

The collective name for the five spectral-weighted irradiances is “ α -opic irradiance.” Each of the five individual α -opic irradiances is named



M

Melanopic Equivalent Daylight Illuminance, Fig. 1 Relative spectral sensitivity curve for the intrinsic photosensitivity of pRGCs for light at the outer surface of the eye (light blue drawn melanopic curve) together with

the curves of rods, S, M, and L cones. Basis: 32-year, 10° standard observer. Bottom figure in logarithmic units. Source: [7]

after its photopigment name, melanopic (pRGC), rhodopic (rod), cyanopic (S cone), chloropic (M cone), and erythropic (L cone) irradiances, as

indicated in Table 1. Often the cone irradiances are simply referred to as S, M, and L cone irradiances.

Melanopic Equivalent Daylight Illuminance, Table 1 Names of the five types of α -opic irradiances and the photopigments after which they are named [8]

Photoreceptor	Photopigment	α -opic irradiance	Symbol	Unit
pRGC	Melanopsin	Melanopic irradiance	Ee, mel	W/m ²
Rod	Rhodopsin	Rhodopic irradiance	Ee, rod	W/m ²
S cone	Photopsin cyanolabe	Cyanopic irradiance	Ee, sc	W/m ²
M cone	Photopsin chlorolabe	Chloropic irradiance	Ee, mc	W/m ²
L cone	Photopsin erythrolabe	Erythropic irradiance	Ee, lc	W/m ²

Figure 2 gives for some typical lamp types the calculated α -opic irradiances for the five photoreceptors, for the condition 1000 lux at the outer surface of the eye. The values are calculated based on the formula given above for the example of melanopic irradiance. For the condition of 1000 lux at the outer surface of the eye the general formula is:

$$E_{e,\alpha} \text{ per } 1000 \text{ lux} = \frac{\int E_e(\lambda) S_\alpha(\lambda) d(\lambda) * 1000 / 683}{\int E_e(\lambda) V(\lambda) d(\lambda)}$$

where

$E_{e,\alpha}$ = α -opic weighted irradiance in W/m²; α to be specified as melanopic, rhodopic, S-cone-, M-cone-, or L-cone-opic

$E_e(\lambda)$ = irradiance for wavelength λ in W/m² (available from the lamp spectrum)

$S_\alpha(\lambda)$ = relative α -opic sensitivity for wavelength λ , where $S_\alpha(\lambda)$ is normalized to the value of 1 at its peak (given in Fig. 1)

$V(\lambda)$ = standardized relative spectral sensitivity curves for photopic vision normalized to the value of 1 at its peak

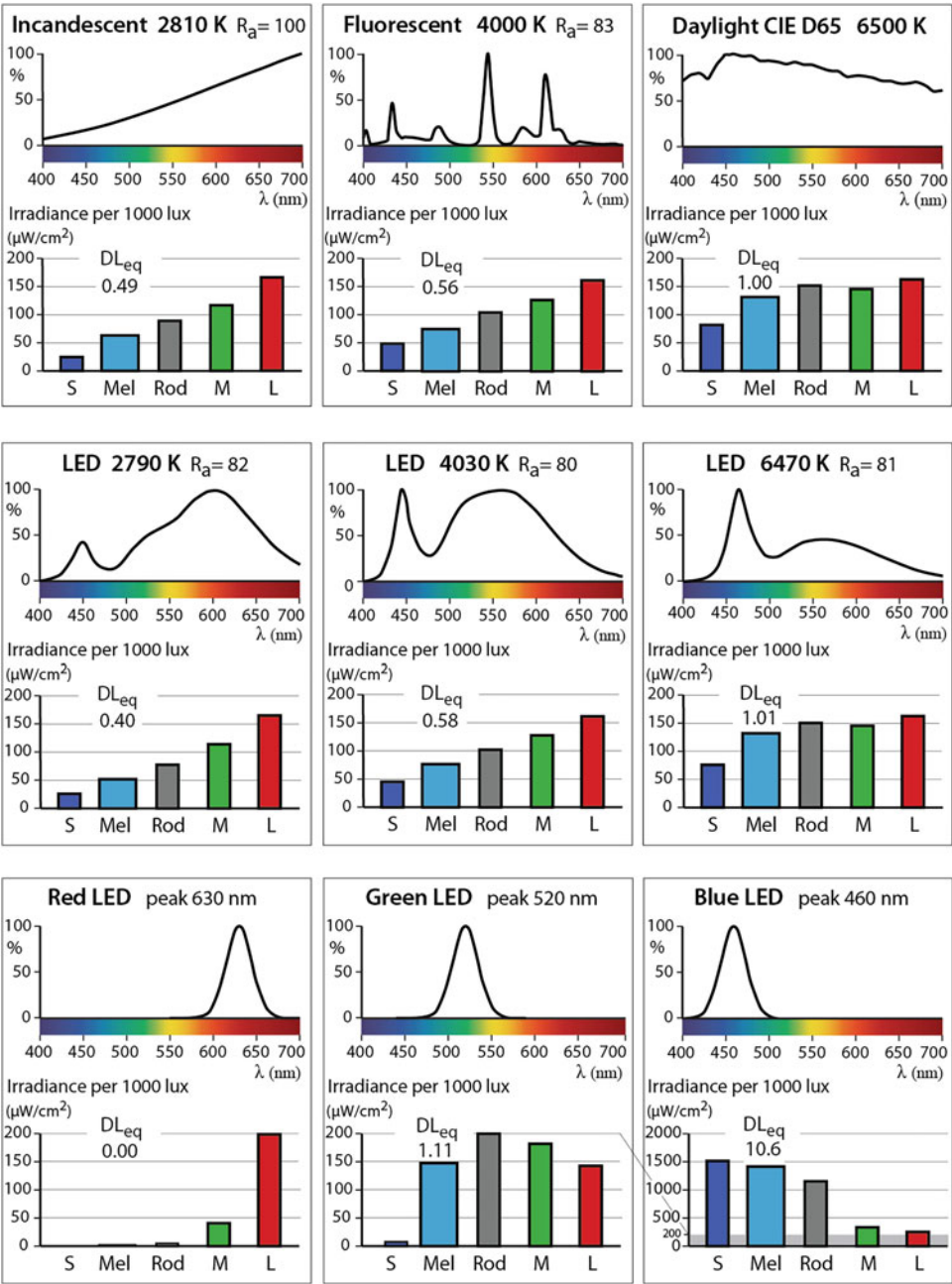
Melanopic Equivalent Daylight Ratio

The five “irradiances per 1000 lux” give a good insight into the effectiveness of a lamp in evoking a reaction in each of the five photoreceptors. For nonvisual biological lighting applications, it may often be interesting to compare the melanopic irradiance of a particular light source with the melanopic irradiance by daylight (of the same lighting level at the eye). For this purpose, Fig. 1

also gives, for each lamp type, the “melanopic equivalent daylight ratio DLeq.” It is the ratio of the melanopic irradiance of the lamp to the melanopic irradiance of 6500 K daylight (CIE standard D65 sky). This ratio is also referred to as melanopic daylight (D65) efficacy ratio [7].

The first two rows of Fig. 2 give the results for white light sources, all with a color rendering Ra of at least 80. The first column shows light sources with a correlated color temperature CCT of ca. 2800 K (incandescent lamp and LED), the second column of ca. 4000 K (fluorescent and LED), and the last column of ca. 6500 K (D65 daylight and LED). When comparing differences in irradiances for the same CCT value, the largest difference occurs for the 2800 K case. Contrary to what is often, wrongly, expected, the incandescent lamp has a larger melanopic irradiance (in this example 63 μ W/cm²) than the LED lamp with a closely same CCT value (52 μ W/cm² or 17% less). The conclusion of this example is in line with other studies [9, 10]. This is of course correct, provided that the illuminance condition at the eye is the same (1000 lux in the above example). When comparing the melanopic irradiances for the different correlated color temperature ranges, it is clear that with higher CCT values, there is a clear trend for higher melanopic irradiances. Popularly phrased: a larger blue content in white light evokes a larger effect in the melanopsin photopigment.

The last row of Fig. 2 shows the results for colored, narrow-spectrum LEDs, respectively for red, green, and blue. Note that the scale of the blue LED is different: It ranges from 0–2000 μ W/cm² while for all other lamps the range is from 0–200 μ W/cm². The blue LED results in much higher irradiances for all photoreceptors. This is,



Melanopic Equivalent Daylight Illuminance, Fig. 2 Alpha-optic irradiances of different lamp types for 1000 lux at the outer surface of the eye. S (S cone) is cyanopic irradiance, Mel (melanopsin) is melanopic irradiance, Rod is rhodopic irradiance, M (M cone) is

chloropic irradiance, and L (L cone) is erythropic irradiance. DL_{eq} is the melanopic equivalent daylight ratio (CIE D65). Bars for melanopic irradiances are shown somewhat broader to make them stand out

in particular, the case for melanopic and cyanopic (S cone) irradiances. The red LED provides no cyanopic and hardly any melanopic irradiance.

These largely contrasting properties of especially red and blue LEDs are important in therapeutic light applications.

Melanopic Equivalent Daylight Illuminance

The melanopic equivalent daylight ratio (D65) of a light source can also be used to express any photometric quantity (as, e.g., luminous flux, luminous intensity, luminance, or illuminance) in a daylight (D65) equivalent quantity. Examples for the 2790 K, 4030 K, and 6470 K white LED lamps of Fig. 2 could look like:

- A 2790 K LED lamp with a melanopic equivalent daylight (D65) ratio of 0.40 and a luminous flux of 500 lumen has a melanopic equivalent daylight (D65) luminous flux of $0.40 * 500 = 200$ lumen.
- A 4030 K LED lighting installation with a melanopic equivalent daylight (D65) ratio of 0.58, providing an illuminance of 600 lux, has a melanopic equivalent daylight (D65) illuminance (Emel,D65eq) of $0.58 * 600 = 348$ lux.
- A 6470 K LED lighting installation with a melanopic equivalent daylight (D65) ratio of 1.01, providing an illuminance of 600 lux, has a melanopic equivalent daylight (D65) illuminance of $1.01 * 600 = 606$ lux.

Cross-References

- [Circadian Rhythms](#)
- [Human-Centric Lighting](#)
- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)
- [Sleep, Alertness, and Light](#)

References

1. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science*. **295**, 1070–1073 (2002)
2. Hankins, M.W., Lucas, R.J.: The primary visual pathway in humans is regulated according to long-term light exposure through the action of a nonclassical photopigment. *Curr. Biol.* **12**(3), 191–198 (2002)
3. Hattar, S., Liao, H.W., Takao, M., Berson, D.M., Yau, K.W.: Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science*. **295**, 1065–1070 (2002)

4. Newman, L.A., Walker, M.T., Brown, R.L., Cronin, T.W., Robinson, P.R.: Melanopsin forms a functional short wavelength photopigment. *Biochemistry*. **42**, 12734–12738 (2003)
5. Brainard, G.C., Hanifin, J.P.: The effects of light on human health and behavior: relevance to architectural lighting. In: CIE x027:2004 Proceedings of CIE Symposium '04 light and health: non-visual effects. pp. 2–16 (2004)
6. Lucas, R.J., Peirson, S.N., Berson, S.N., Brown, T.M., Cooper, H.M., Czeisler, C.A., Figueiro, M.G., Gamlin, P.D., Lockley, S.W., O'Hagan, J.B., Price, L.L.A., Provencio, I., Skene, D.J., Brainard, G.C.: Measuring and using light in the melanopsin age. *Trends Neurosci.* **37**(1), 1–9 (2014)
7. CIE: CIE International Standard CIE 026:2018 CIE system for metrology of optical radiation for ipRGC-influenced responses to light. Vienna (2018)
8. CIE: CIE Technical Note 003:2015, Report on the first international workshop on circadian and neurophysiological photometry, 2013 (2015)
9. Van Bommel, W.J.M.: Incandescent replacement lamps and health. *LED Prof. Rev.* **19**, 32–35 (2010)
10. Kobav, M.B., Bizjak, G.: LED spectra and melatonin suppression action function. *Light Eng.* **20**(3), 15–22 (2012)

Melanopsin Contributions to Human Brightness Perception

Sei-ichi Tsujimura¹ and Yoshika Takahashi²

¹Graduate School of Design and Architecture, Nagoya City University, Aichi, Japan

²Graduate School of Engineering, Chiba University, Chiba, Japan

Synonyms

[Brightness perception based on physiology; Contribution of melanopsin-containing retinal ganglion cell to brightness perception; Brightness perception associated with photoreceptors' stimulations](#)

Definition

The contribution of melanopsin to brightness perception in humans is the role of photopigment melanopsin or of melanopsin-containing retinal

ganglion cell in brightness perception. Based on photometry, it is assumed that rods and cones are the only photoreceptors responsible for brightness perception. However, investigation of brightness perception should be reassessed based on stimulations of all photoreceptor classes including rods, cones, and melanopsin-containing retinal ganglion cells.

Overview

The discovery of melanopsin-containing retinal ganglion cells has led to a fundamental reassessment of visual and nonvisual processes responsible for encoding ambient illuminance. This entry provides a brief introduction of the functional role driven by melanopsin-containing retinal ganglion cells in brightness perception. It also discusses the practical difficulties involved in investigating this subject and outlines potential avenues for future research, which would benefit scientists and engineers who wish to study this topic.

Discovery of Melanopsin-Containing Retinal Ganglion Cells

Humans do not perceive light, but rather see objects by reflection. Light is characterized as a medium that conveys basic visual attributes such as color and luminance from the environment to our eye. The discovery of melanopsin-containing retinal ganglion cells (hereafter, “melanopsin cells”) has led to a fundamental reassessment of nonimage forming visual processing, such as circadian photo-entrainment and the pupillary light reflex [1]. This suggests that melanopsin cells play an important role in encoding ambient illuminance [1, 2]. Signals from melanopsin cells, stimulated by environmental light, contribute to nonimage forming processing along with signals from the classical photoreceptors viz. rods and cones.

Growing evidence indicates that melanopsin cells contribute not only to nonimage forming processing, but also to visual perception [2, 3]. For example, mice lacking classical photoreceptors but having melanopsin cells as the only

functional photoreceptor can discriminate grating patterns. Similarly, a blind person lacking functional classical photoreceptors can detect a light step. The apparent brightness is enhanced as stimulation of melanopsin increases in humans [3–5]. Although these results support the fact that melanopsin cells play an important role in the conventional visual pathway in humans, their functional role remains unclear.

Difficulties in Estimating the Contribution of Melanopsin to Brightness Perception

In humans, it is difficult to investigate whether melanopsin cells contribute to visual attributes. This limitation stems primarily from the challenges in selectively stimulating melanopsin cells while avoiding other photoreceptors and post-receptor mechanisms. For example, when brightness is measured as a function of the level of light source emission its perception can be explained by consolidated increases in stimulation of the classical photoreceptors and melanopsin cells. Although it is possible to use a monochromatic light stimulus with an approximate wavelength of 480 nm to selectively stimulate melanopsin cells, the classical photoreceptors would also be strongly stimulated, inducing luminance and color perception.

Here, therefore, to stimulate melanopsin independently of other photoreceptors particular light stimuli known as metamers were used. Metamers have the same color and luminance but have distinct spectral power distribution. These stimuli can independently stimulate melanopsin while keeping other photoreceptors silent, ensuring that the color and luminance of these stimuli do not change. The way of stimulation is well-known as silent-substitution paradigm in psychophysics [6].

Melanopsin and Human Brightness Perception

The term “brightness” is used to denote the perceived apparent intensity of ambient light and is

different from the photometric term “luminance.” Luminance is estimated from luminous efficiency functions, that is, a linear combination of the L- and M-cone spectral sensitivities measured by heterochromatic flicker photometry at high temporal frequency, without considering any input from the S-cones and melanopsin cells. Brightness does not correlate with photometric luminance based on the assumptions of the conventional rod/cone visual pathway [7]. Furthermore, brightness does not require spatial information [8] and increases with correlated color temperature even for isoluminant stimuli [9]. Here, a brightness-matching experiment for assessing melanopsin contribution to brightness perception will be introduced.

Brightness-Matching Experiment

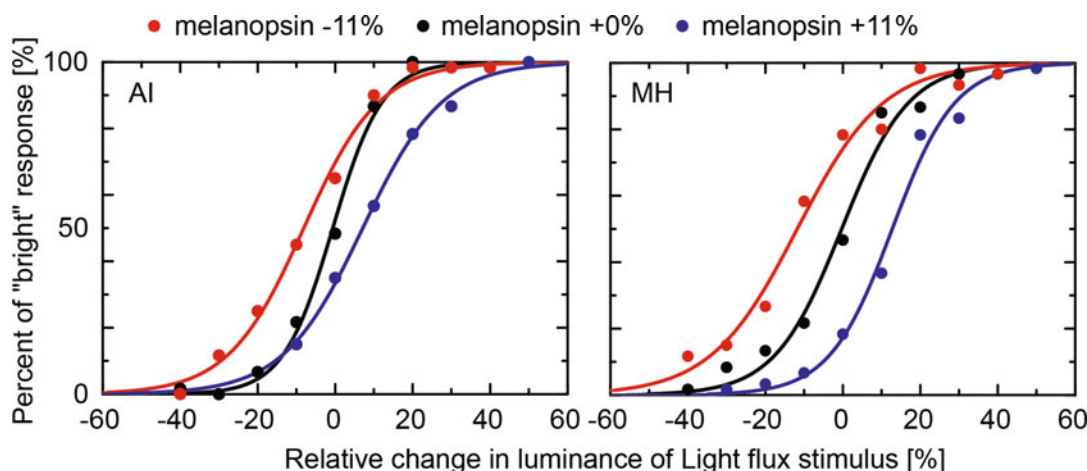
A test stimulus and reference stimulus were sequentially presented to subjects, separated by audible tones. The reference stimulus was presented after the test stimulus and vice versa. The background transitioned over 2 s to the first stimulus, which was presented for 1 s. Then, the first stimulus transitioned over 2 s to the second stimulus, which was presented for 1 s. Finally, the second stimulus reverted over 2 s to the background. This was done to minimize the effect of abrupt changes in stimulus. The observers had to choose between the first and second stimulus

presentation with regard to brightness. Luminance of the nine different light flux stimuli varied between -40% and $+50\%$ relative to the background.

The test stimulus subtended 95° of the visual angle. Observers were seated 4.5 cm in front of the diffuser. Each session lasted approximately 40 min, including an initial adaptation of 5 min, during which time observers were instructed to keep their right eye open, especially when the integrating sphere provided test and reference stimuli for 10-s periods. The left eye was masked throughout the test. Data were averaged from three or five sessions.

In the experiment, brightness was matched between the melanopsin stimulus and the light flux stimulus by adjusting the luminance of the light flux stimulus. A two-interval alternative forced-choice procedure with the method of constant stimuli was used to identify a light flux stimulus with brightness matched to the melanopsin stimulus. Three melanopsin stimuli (-11% , 0% , and $+11\%$) were used as test stimuli. Because only melanopsin stimulation of the test stimulus varied from the background, the apparent brightness could therefore be attributed to the difference in stimulation of melanopsin.

The brightness of the melanopsin stimulus was compared with the light flux stimuli (Fig. 1). The



Melanopsin Contributions to Human Brightness Perception, Fig. 1 Brightness matching. Percentage of responses when the observer perceived that the light flux

stimulus was brighter than the melanopsin stimulus. Psychometric functions were fitted to the data from the logistic function

curve represents a logistic function, and a horizontal shift of the curve from zero represents the amount of apparent brightness induced. Results show that the curve shifted to the right as melanopsin stimulation increased, indicating that the observers perceived increased brightness.

When the shift was compared among conditions, it was consistently observed that shifts for +11% melanopsin stimulation were significantly larger than those for -11% and 0% melanopsin stimulation in all observers. Moreover, the shifts for -11% melanopsin stimulation were significantly smaller than those for 0% melanopsin stimulation in all observers (Fig. 2, open symbols). These results indicate that all observers perceived progressively more brightness as melanopsin stimulation increased.

Flicker-Photometry Experiment

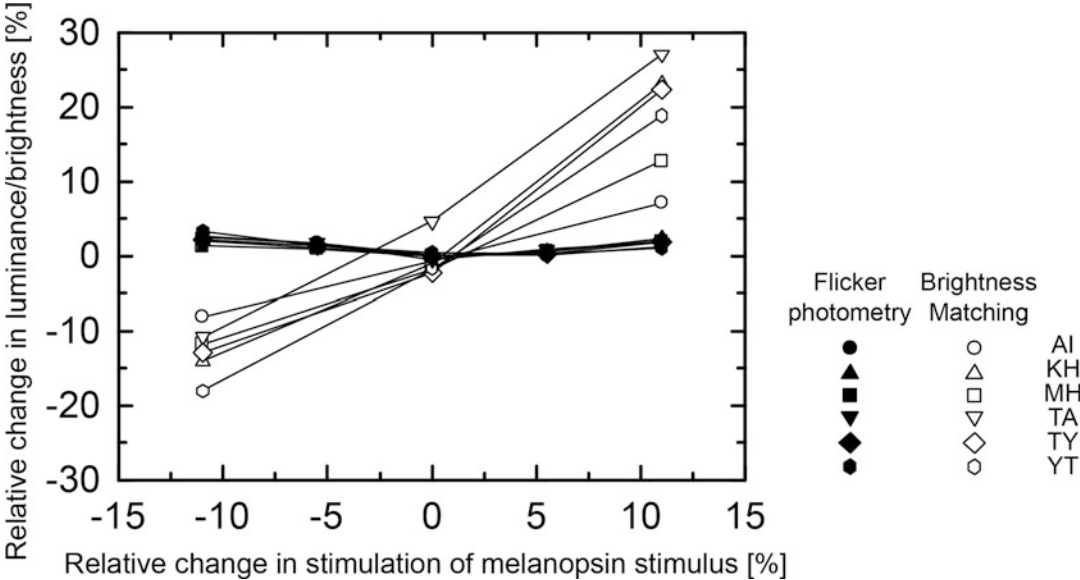
Individual variations in luminous efficiency function among observers were examined. The luminance of the melanopsin stimulus was measured by flicker photometry with the method of adjustment in all observers (Fig. 2, solid symbols). The relative change in luminance for five melanopsin

stimuli ranged from -0.5% to 2.4%, with an average of 1.2%. No significant difference in the slope of the linear regression line was found for luminance by flicker photometry. These results show that the luminance variation for observers was negligible; this indicates that the melanopsin did not contribute to luminance perception.

Spectral Sensitivity Curve for Melanopsin Photoreception

Test stimuli were represented in a receptor-stimulation space that used stimulations of three types of cones and melanopsin. The receptor-stimulation space is a natural extension of cone-stimulation space (also called cone-excitation space). Because light stimulation correlates with the cell's stimulation, not with the cell's activity a term "stimulation" is used here. Cone-stimulation space uses three fundamentals, corresponding to the stimulation of the three kinds of retinal cones. The fundamentals were so designed that the total amount of stimulation of L and M cones was equivalent to the photopic luminous efficacy

M



Melanopsin Contributions to Human Brightness Perception, Fig. 2 Relative change in brightness (open symbols) and in luminance (solid symbols) for all observers. The shifts for the +11% melanopsin stimulation condition

were significantly greater than those for the -11% and 0% conditions while the shifts for the -11% condition were significantly smaller than those for the 0% condition

function $V(\lambda)$. In addition to these fundamentals, a spectral sensitivity curve for melanopsin was used. These were mapped onto four orthogonal axes in receptor-stimulation space.

The spectral sensitivity of photopigment melanopsin at cornea was estimated using a pigment template nomogram, a standard shape of visual pigment absorption coefficient, $S(\lambda)$ [10], with a peak wavelength ($\lambda_{\max}$) of 480 nm. The lens and macular pigment density spectra, $L(\lambda)$ and $M(\lambda)$, in the peripheral visual field were those used by Stockman et al. [11]. The fraction of incident light absorbed by the receptor depends on the peak axial optical density (D_{peak}). As there was no prior reference, D_{peak} was kept as 0.1. Hence, the spectral sensitivity curve for melanopsin photoreception, $\text{Mel}(\lambda)$, was calculated based on light absorption by the lens and macular pigment and photopigment peak axial optical density [12, 11].

$$\text{Mel}(\lambda) = A(\lambda) \left(1 - 10^{-D_{\text{peak}} \cdot S(\lambda)} \right)$$

where

$$A(\lambda) = 10^{-(L(\lambda) + M(\lambda))}$$

Although short-wavelength sensitive cones (S-cones) and melanopsin photoreception do not affect the photopic luminance efficiency function (i.e., luminance), photopic luminance units (cd m^{-2}) were used. Similar to S-cone stimulation [13], the melanopsin stimulation was defined as the amount of stimulation produced by an equal energy spectrum of luminance 1.0 cd m^{-2} . The resultant spectral sensitivity function of melanopsin in melanopsin cell in the peripheral visual field displayed a peak of 872.1 (i.e., K_m) at a peak wavelength of 493 nm. The shape of this spectral sensitivity curve was similar to that proposed by Lucas et al. [14]. Tsujimura et al. further considered the human macular pigment density at 10-deg for the estimation [15]. Although the shapes of spectral sensitivity curves are slightly different from that proposed by Lucas et al., the photoreceptor stimulations

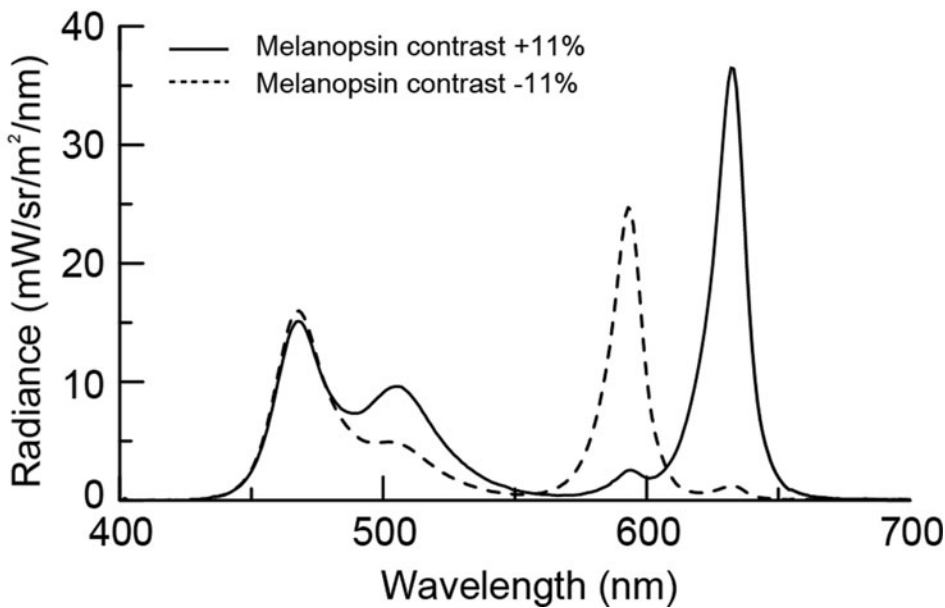
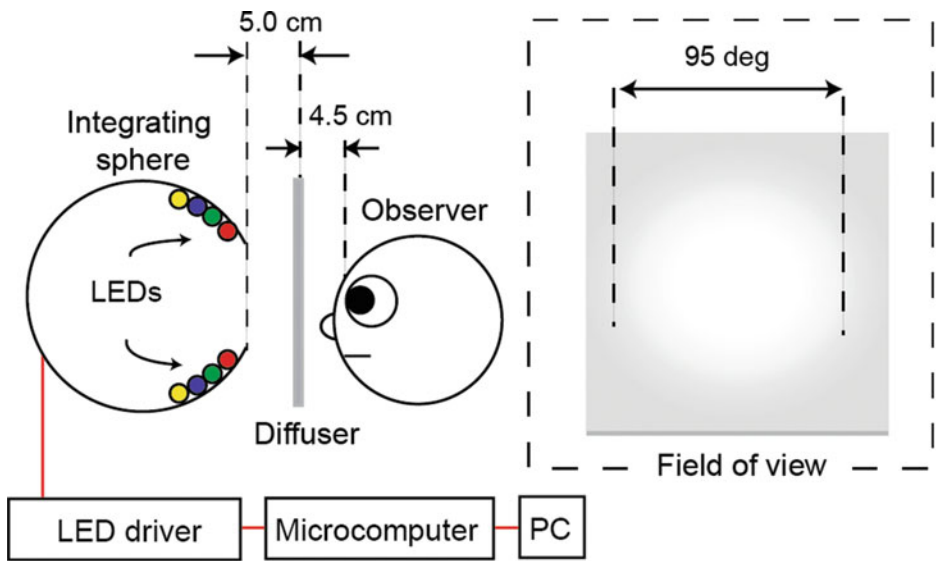
calculated from the primaries were in agreement with the experiment.

Test Stimuli in Silent-Substitution Paradigm for Independent Stimulation of Melanopsin

To stimulate melanopsin independently of other photoreceptors the silent-substitution paradigm was used [6]. A four-primary illumination system was used to investigate melanopsin-cone interactions [2]. Two test stimuli were presented: (i) melanopsin stimulus with varying melanopsin stimulation while keeping stimulation of the three types of cones constant (Fig. 3), and (ii) light flux stimulus with varying radiant flux of the stimuli with constant spectral composition which reduced/increased the radiant flux uniformly at all wavelengths. (See published literature for detailed description of the system [3, 15].)

Test stimuli were represented in a receptor-stimulation space that uses stimulations of three types of cones and melanopsin [15]. The background stimulus had a CIE coordinate of (0.351, 0.311) and a luminance of 356 cd m^{-2} . The receptor stimulations for the background were 255 cd m^{-2} , 102 cd m^{-2} , and 379 cd m^{-2} for the L-, M-, and S-cone, respectively, and 1159 cd m^{-2} for melanopsin. The ratio of L- to M-cone stimulation was 2.5, identical to that for 570 nm monochromatic light. For a 570 nm yellow background, the red-green ratio in flicker detection was almost constant with respect to changes in mean luminance. Retinal illuminance in response to the background was 2.8 log photopic troland and 3.2 scotopic troland with a pupil size of 1.5 mm to minimize the involvement of rods. The retinal irradiance in the photon flux of the control stimulus was $15.1 \text{ log photons cm}^{-2} \text{ s}^{-1}$.

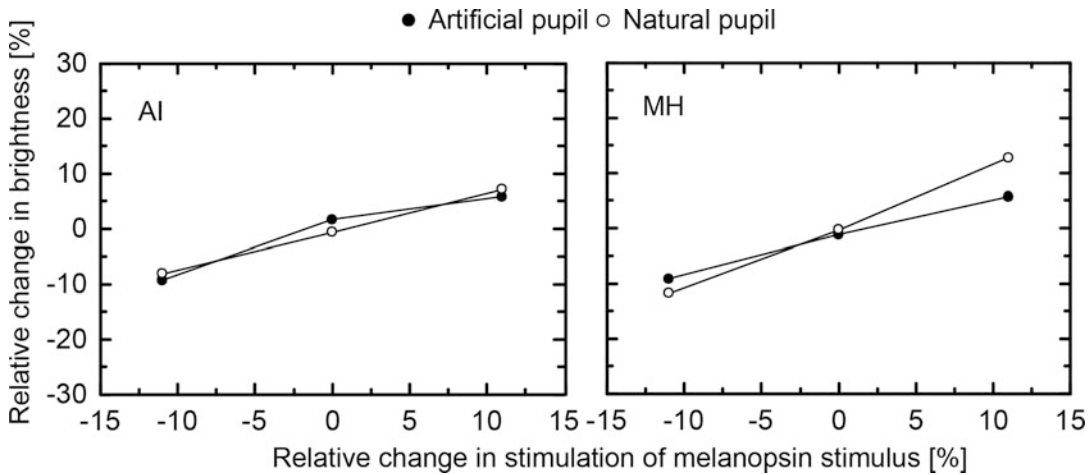
In general, the silent-substitution paradigm is a sophisticated technique to investigate the function of melanopsin photoreceptors. However, it must be noted that one of limitations of the silent substitution is that the change in stimulation (i.e., contrast) of target photoreceptor is small. The change in independent stimulation of melanopsin



Melanopsin Contributions to Human Brightness Perception, Fig. 3 Experimental setup. The apparatus consists of an integrating sphere controlled by microcomputer (left), and the spectra used for high- and low-melanopsin stimuli (right)

strongly depends on the choice of primaries of light source and importantly the change is usually not so large. This limitation makes it difficult to investigate a function associated with melanopsin stimulation. Furthermore, because the silent-substitution paradigm is based on spectral

sensitivity curves of photoreceptors results are sometimes contaminated from the nontarget photoreceptors. In addition, spectral sensitivity curves vary among individuals according to optical properties at retina. Therefore, the experiment must be conducted carefully.



Melanopsin Contributions to Human Brightness Perception, Fig. 4 Relative changes in brightness with and without artificial pupil

Brightness-Matching Experiment with Artificial Pupil

As melanopsin cells contribute greatly to the pupillary pathway, results of the brightness-matching experiment may be influenced by pupil size. Therefore, to control for the pupil size the above experiment using an artificial pupil of 1.5 mm diameter was repeated. Results were similar to those of the previous experiment (Fig. 4), indicating that the influence of variation in pupil size was negligible.

Luminance, Color, Brightness, and Melanopsin in Photometry

In photometry, the apparent brightness of a light source does not correlate significantly with its luminance, especially when a chromatic stimulus is used. Therefore, it was assumed that color signals contribute to brightness. For example, a stimulus appears brighter as chromatic saturation increases (sometimes referred to as the Helmholtz–Kohlrausch effect), which explains the contribution of color to brightness.

Brightness is independent of photometric luminance based on the assumption of the

conventional rod/cone visual pathway [7]. Furthermore, brightness perception does not require spatial information [8] and increases with color temperature, even for isoluminant stimuli [9]. Because previous studies did not consider melanopsin cells, it is likely that a part of these phenomena can be explained by a melanopsin contribution to brightness perception.

Natural light has complex spectra and the spectral sensitivity curves of the photoreceptors overlap. Therefore, the natural light usually stimulates all photoreceptor classes. To evaluate brightness perception for natural light based on physiological substrate, it is important to estimate the weights of combination of the photoreceptors. To achieve this, color matching functions based on all photoreceptor classes including melanopsin photoreceptors might be required. Further research on the integration process of melanopsin and the other photoreceptors will hopefully provide an answer to this question in photometry.

Cross-References

- [Cone Fundamentals, Stockman-Sharpe](#)
- [Melanopsin Retinal Ganglion Cells](#)

References

1. Lucas, R.J., Douglas, R.H., Foster, R.G.: Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nat. Neurosci.* **4**(6), 621–626 (2001). <https://doi.org/10.1038/88443>
2. Dacey, D.M., Liao, H.W., Peterson, B.B., Robinson, F.R., Smith, V.C., Pokorny, J., Yau, K.W., Gamlin, P.D.: Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature*. **433**(7027), 749–754 (2005). <https://doi.org/10.1038/nature03387>
3. Brown, T.M., Tsujimura, S., Allen, A.E., Wynne, J., Bedford, R., Vickery, G., Vugler, A., Lucas, R.J.: Melanopsin-based brightness discrimination in mice and humans. *Curr. Biol.* **22**(12), 1134–1141 (2012). <https://doi.org/10.1016/j.cub.2012.04.039>
4. Zele, A.J., Adhikari, P., Feigl, B., Cao, D.: Cone and melanopsin contributions to human brightness estimation: reply. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **35**(10), 1783 (2018). <https://doi.org/10.1364/JOSAA.35.001783>
5. Yamakawa, M., Tsujimura, S.I., Okajima, K.: A quantitative analysis of the contribution of melanopsin to brightness perception. *Sci. Rep.* **9**(1), 7568 (2019). <https://doi.org/10.1038/s41598-019-44035-3>
6. Donner, K.O., Rushton, W.A.: Retinal stimulation by light substitution. *J. Physiol.* **149**, 288–302 (1959)
7. Rea, M., Bullough, J., Freyssinier-Nova, J., Bierman, A.: A proposed unified system of photometry. *Light. Res. Technol.* **36**(2), 85–109 (2004). <https://doi.org/10.1191/1365782804li114oa>
8. Barlow Jr., R.B., Verrillo, R.T.: Brightness sensation in a ganzfeld. *Vis. Res.* **16**(11), 1291–1297 (1976)
9. Berman, S., Jewett, D., Fein, G., Saika, G., Ashford, F.: Photopic luminance does not always predict perceived room brightness. *Light. Res. Technol.* **22**(1), 37–41 (1990)
10. Dartnall, H.J.: The interpretation of spectral sensitivity curves. *Br. Med. Bull.* **9**(1), 24–30 (1953)
11. Stockman, A., Sharpe, L.T., Fach, C.: The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vis. Res.* **39**(17), 2901–2927 (1999)
12. Baylor, D.A., Nunn, B.J., Schnapf, J.L.: Spectral sensitivity of cones of the monkey *Macaca fascicularis*. *J. Physiol.* **390**, 145–160 (1987)
13. Boynton, R.M., Kambe, N.: Chromatic difference steps of moderate size measured along theoretically critical axes. *Color Res. Appl.* **5**(1), 13–23 (1980)
14. Lucas, R.J., Peirson, S.N., Berson, D.M., Brown, T.M., Cooper, H.M., Czeisler, C.A., Figueiro, M.G., Gamlin, P.D., Lockley, S.W., O'Hagan, J.B.: Measuring and using light in the melanopsin age. *Trends Neurosci.* **37**(1), 1–9 (2014)
15. Tsujimura, S., Ukai, K., Ohama, D., Nuruki, A., Yunokuchi, K.: Contribution of human melanopsin retinal ganglion cells to steady-state pupil responses. *Proc. Biol. Sci.* **277**(1693), 2485–2492 (2010). <https://doi.org/10.1098/rspb.2010.0330>

Melanopsin Retinal Ganglion Cells

Robert Lucas

Faculty of Biology Medicine and Health,
University of Manchester, Manchester, UK

Synonyms

Intrinsically photosensitive retinal ganglion cells; ipRGCs; mRGCs; Photosensitive retinal ganglion cells; pRGCs

Definition

A subset of retinal ganglion cells are characterized by their expression of melanopsin and their ability to respond directly to light. Melanopsin is an opsin protein structurally and phylogenetically related to rod and cone opsins. It binds retinaldehyde as a chromophore and activates a G-protein signaling cascade upon photon absorption. As a result, melanopsin retinal ganglion cells are depolarized by light even when physically, pharmacologically, or genetically isolated from rod and cone influences. This allows them to function as an independent origin of visual information. mRGCs are known to encode environmental brightness for such visual reflexes as circadian photoentrainment and regulating pupil size. They likely also make as yet incompletely defined contributions to other visual processes including perceptual vision.

Discovery and Functions

Until around the turn of the millennium it was considered self-evident that all aspects of human vision could be ultimately attributed to the activity of rod and/or cone photoreceptors. With hindsight, it is possible to find isolated examples of findings in several branches of vision science that questioned this assumption, but the concerted effort that led to the discovery of a third photoreceptor type came from researchers interested

in a particular visual function – circadian entrainment [1].

Like almost all life on earth, humans have an internal circadian clock that oscillates with a period close to 24 h. In order to fulfill its function of coordinating physiology and behavior to the varying demands of the astronomical day, this clock needs to be accurately set (or “entrained”) to local time. Entrainment relies upon a representation of the light:dark cycle provided by a dedicated retinal projection to the hypothalamic site of the circadian oscillator, the suprachiasmatic nuclei (or SCN). Studies of human subjects and rodent models of retinal dystrophy revealing that circadian entrainment was buffered against massive loss of rods and cones culminated in the demonstration that this function was retained even when these conventional photoreceptors were completely absent [2]. An explanation for this finding was provided by the discovery that the retinal ganglion cells projecting to the SCN are in fact directly photoreceptive [3].

Melanopsin is one of a number of opsin proteins known to be expressed outside of the eyes in nonmammalian vertebrates. It was initially discovered in the photosensitive dermal melanophores of *Xenopus laevis* [4]. Orthologues were subsequently identified in mammalian genomes and found to be expressed in the retinal ganglion cells innervating the SCN. These ganglion cells were shown to lack their intrinsic photosensitivity in mice lacking melanopsin [5], while inclusion of melanopsin was shown to be sufficient to induce photosensitivity in a variety of non-light-responsive cell types [6–8].

Although initially discovered in the search to understand circadian entrainment, melanopsin retinal ganglion cells are now thought to contribute to a wide variety of visual responses [9]. mRGCs project not just to the SCN but also to other parts of the hypothalamus, the pretectum, the superior colliculus, the visual thalamus, and the nucleus of the optic tract [10, 11]. Their established functions extend to providing the pupil light reflex, regulating sleep/alertness, and suppressing pineal melatonin production. There is also evidence that they regulate mood and retinal development and

influence retinal physiology by signaling to other cells in the retina.

At present, the nature and extent of mRGC contributions to perceptual vision is unclear. The appearance of mRGC fibers in projections to the dorsal lateral geniculate nucleus and established functional connections to other cell types in the retina via synaptic and gap junction connections raise the possibility that mRGCs could have both direct and modulatory influences on perceptual vision. Evidence from both mice and primates supports the view that melanopsin can both provide an origin for aspects of visual light adaptation and itself support visual discrimination [12–15]. In primates, at least some ipRGCs receive antagonistic input from short- and long/medium-wavelength cones and may contribute to blue:yellow discrimination [16].

Anatomy and Physiology

The defining feature of mRGCs is their expression of melanopsin, and consequent intrinsic photosensitivity. In rodents, at least six morphologically distinct classes of retinal ganglion cell (termed M1 to 6) meet this criterion [17]. The best characterized of these is the M1 class, which has the highest melanopsin expression (and the most prominent intrinsic light response) and is responsible for both circadian entrainment and the pupil light reflex. Other classes differ in their central projections, the nature and spatial extent of dendritic arborization, and their degree of melanopsin expression [17]. Morphological diversity is also a characteristic of the mRGC population in primates and humans, although fewer subtypes have so far been identified in primate than rodents [16, 18].

The dendrites of mRGCs are sites of melanopsin-driven phototransduction but also of synaptic inputs from bipolar and amacrine cells. As a result, in the intact retina, the spike-firing pattern of mRGCs is defined in part by melanopsin photoreception and in part by synaptic influences originating in signals from rods and cones. As mRGCs lack the membranous disks used by rods and cones to maximize

photopigment concentration, they have very poor photon capture efficiency. As a result, melanopsin has an appreciable impact on mRGC firing only at rather high light intensities (several decades above the threshold for cone activation). Even at these light levels, the rate of melanopsin photoisomerization is low and its phototransduction cascade has to have high gain [19]. This is achieved by having a long activation time, with the consequence that the melanopsin light response has poor temporal resolution. Under some circumstances, this phenomenon can be extreme with light-induced firing of mRGCs and melanopsin-driven pupil constriction lasting several tens of seconds after the light has been turned off. Under more natural conditions, these sensory characteristics imply that melanopsin provides a low spatiotemporal resolution signal of environmental brightness, with mRGCs relying upon input from rods and cones to encode lower light levels and to track higher frequency changes in light intensity [20].

Melanopsin, like all animal opsins, is a member of the G-protein coupled receptor superfamily of proteins. The molecular identity of the critical elements of its phototransduction cascade is unknown. Melanopsin can signal via a G-protein of the Gq/11 class, which links opsin activation to membrane depolarization via phospholipase C and TRP channels [21]. However, melanopsin also likely interacts with other signaling cascades in at least some mRGCs [22, 23].

Spectral Sensitivity

Measuring the spectral absorbance properties of melanopsin directly has proved challenging. It is not found in sufficient concentrations to allow direct *in vivo* microspectrophotometry. While it has been possible to purify melanopsin from harvested retinas and from heterologous cell-based expression systems, the quality of these preparations falls far short of that achieved with rod or cone opsins. As a result, a precise description of melanopsin's spectral absorbance in the dark state, and following light absorption, remains

elusive. Nonetheless, action spectra for melanopsin-driven activation of mRGC activation and downstream responses in a variety of mammalian species are consistent with the hypothesis that in the dark state melanopsin's spectral sensitivity can be well described by the standard opsin: vitamin A1 template peaking around 480 nm [20]. This puts it close to the spectral sensitivity of human rod opsin but far from that of the cone photoreceptors that are conventionally considered to dominate vision at moderate and high light levels. Importantly, its spectral sensitivity is also significantly short wavelength shifted compared to $V(\lambda)$ (the spectral efficiency function of photopic vision), which is used in quantifying illuminance. This implies that spectrally distinct natural and artificial light sources might have quite different effective intensity for melanopsin even when matched for illuminance [20] and has led the Committee International de l'Eclairage to establish an independent measurement standard for melanopsin-dependent light responses (CIE S 026/E:2018).

Cross-References

- [Architectural Lighting](#)
- [Ganglion Cells](#)
- [Melanopic Equivalent Daylight Illuminance](#)
- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)
- [Sleep, Alertness, and Light](#)

References

1. Bailes, H.J., Lucas, R.J.: Melanopsin and inner retinal photoreception. *Cell. Mol. Life Sci.* **67**, 99–111 (2010)
2. Foster, R.G., Hankins, M.W.: Non-rod, non-cone photoreception in the vertebrates. *Prog. Retin. Eye Res.* **21**, 507–527 (2002)
3. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science*. **295**, 1070–1073 (2002)
4. Provencio, I., Jiang, G., De Grip, W.J., Hayes, W.P., Rollag, M.D.: Melanopsin: an opsin in melanophores, brain, and eye. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 340–345 (1998)

5. Lucas, R.J., et al.: Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science*. **299**, 245–247 (2003)
6. Melyan, Z., Tarttelin, E.E., Bellingham, J., Lucas, R. J., Hankins, M.W.: Addition of human melanopsin renders mammalian cells photoresponsive. *Nature*. **433**, 741–745 (2005)
7. Qiu, X., et al.: Induction of photosensitivity by heterologous expression of melanopsin. *Nature*. **433**, 745–749 (2005)
8. Panda, S., et al.: Illumination of the melanopsin signaling pathway. *Science*. **307**, 600–604 (2005)
9. Schmidt, T.M., et al.: Melanopsin-positive intrinsically photosensitive retinal ganglion cells: from form to function. *J. Neurosci*. **31**, 16094–16101 (2011)
10. Ecker, J.L., et al.: Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron*. **67**, 49–60 (2010)
11. Brown, T.M., et al.: Melanopsin contributions to irradiance coding in the thalamo-cortical visual system. *PLoS Biol*. **8**, e1000558 (2010)
12. Allen, A.E., Martial, F.P., Lucas, R.J.: Form vision from melanopsin in humans. *Nat. Commun*. **10**(1), 2274 (2019)
13. Allen, A.E., et al.: Melanopsin-driven light adaptation in mouse vision. *Curr. Biol*. **24**(21), 2481–2490 (2014)
14. Zele, A.J., et al.: Cone and melanopsin contributions to human brightness estimation. *J. Opt. Soc. Am. A Opt. Image Sci. Vis*. **35**(4), B19–B25 (2018)
15. Spitschan, M., et al.: The human visual cortex response to melanopsin-directed stimulation is accompanied by a distinct perceptual experience. *Proc. Natl. Acad. Sci. U. S. A.* **114**(46), 12291–12296 (2017)
16. Dacey, D.M., et al.: Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature*. **433**, 749–754 (2005)
17. Sand, A., Schmidt, T.M., Kofuji, P.: Diverse types of ganglion cell photoreceptors in the mammalian retina. *Prog. Retin. Eye Res*. **31**, 287–302 (2012)
18. Liao, H.W., et al.: Melanopsin-expressing ganglion cells on macaque and human retinas form two morphologically distinct populations. *J. Comp. Neurol*. **524**(14), 2845–2872 (2016)
19. Do, M.T., Yau, K.W.: Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev*. **90**, 1547–1581 (2011)
20. Lucas, R.J., et al.: Measuring and using light in the melanopsin age. *Trends Neurosci*. **37**, 1–9 (2014)
21. Hughes, S., Hankins, M.W., Foster, R.G., Peirson, S. N.: Melanopsin phototransduction: slowly emerging from the dark. *Prog. Brain Res*. **199**, 19–40 (2012)
22. Jiang, Z., et al.: Cyclic-nucleotide- and HCN-channel-mediated phototransduction in intrinsically photosensitive retinal ganglion cells. *Cell*. **175**(3), 652–664.e12 (2018)
23. Sonoda, T., et al.: Melanopsin phototransduction is repurposed by ipRGC subtypes to shape the function of distinct visual circuits. *Neuron*. **99**(4), 754–767.e4 (2018)

Memory Color

Christoph Witzel¹ and Karl Gegenfurtner²

¹School of Psychology, University of Southampton, Southampton, UK

²Department of Psychology, Giessen University, Giessen, Germany

Synonyms

Canonical color; Diagnostic color; Typical color

Definition

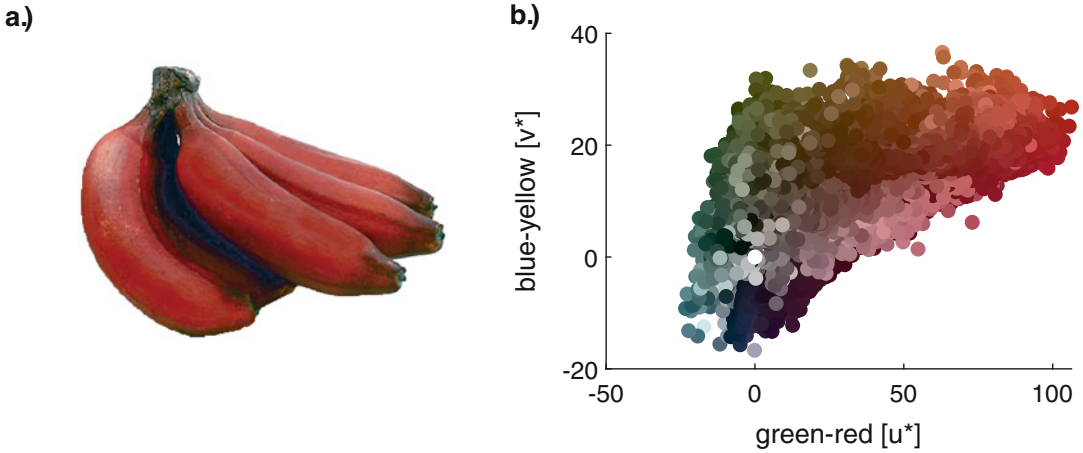
A memory color is the color a beholder considers to be characteristic for an object based on their experience with that object. For example, the memory color of a banana is yellow for most people because they associate a banana with yellow in their memory.

Conceptual Clarifications

The concept of memory color should not be confused with color memory. Color memory is a general term that describes an observer's ability to memorize a color. In contrast, a memory color is an association between a color and an object in memory (for review, see [1, 2]). The notion of memory color is strongly related to *memory color effects*, *canonical and typical colors*, and *color diagnosticity*; but a closer look reveals that there are also important differences.

Memory Color Effects

The term “memory color” is historically linked to the idea of a memory color effect (for review, see [2]). The term has been coined by Ewald Hering in 1878, who associated memory colors with the idea that knowledge about the typical colors of objects affects the perception of the actual color of the objects. However, the fact that an observer knows about typical object colors does not necessarily imply that this knowledge influences the



Memory Color, Fig. 1 What is the banana’s memory color? Panel **a** shows a red Dacca banana to illustrate the fact that a memory color, such as the yellow of a banana, depends on the observer’s familiarity with the object, e.g., with yellow Cavendish rather than red Dacca bananas. Panel **b** illustrates the color distribution that corresponds

to the photo in panel **a**. X- and Y-axes are the green-red (u^*) and blue-yellow (v^*) dimensions of CIELUV color space. The photo also varies in Lightness (L^*), the third dimension, which is not shown to avoid clutter. Photo taken from https://upload.wikimedia.org/wikipedia/ar/b/b4/Red_banana.jpg

way they see the actual object colors. For this reason, it is sensible to disentangle the concept of a memory color from the idea of a memory color effect on color appearance (as, e.g., in [3]), and hence, not all research on memory colors is about the memory color effect (for reviews, see [1, 4–6]).

Canonical and Typical Colors

The term “memory color” ultimately refers to the same phenomenon as “canonical” (e.g., in [7–9]) and “typical color” (e.g., in [5, 10]), though with different emphases and connotations. The idea of a “canonical color” suggests that there is a standard (or reference) that is shared among all observers and that unambiguously determines the association between an object and its canonical color. A typical color focuses on the color as a property of an object. Hence, the ideas of “typicality” or “prototype” refer primarily to the occurrence of object colors in the physical environment.

In contrast, the concept of memory color emphasizes the psychological, subjective nature of perception and memorization. There are two major aspects underlying the subjective nature of memory colors. First, memory colors depend on the observer’s familiarity and experience with an

object. For example, not all bananas are yellow like the most common type of Cavendish bananas. They can also be red, such as the Red Dacca banana (cf. Fig. 1a). Nevertheless, most people associate a banana with yellow, not red, because they are more familiar with yellow bananas. Second, memory colors depend on which color an observer considers to be representative. The colors of most objects correspond to complex color distributions (cf. Fig. 1b), which vary across exemplars and over time. Even the yellow Cavendish bananas can be green or brown depending on their ripeness. Although we know this, we consider the yellow color of a ripe banana to be the most representative and typical.

At the same time, the ideas of canonical and typical colors also require, at least implicitly, assumptions about a beholder (e.g., [11]). Often an experimenter determines which color is canonical or typical for an object, or a computer program computes typicality based on image statistics. In these cases, the choice of canonical or typical colors depends either on the experimenter’s prior assumptions of objects and their criteria of typicality, or on the sample of images and the criteria of typicality processed by the algorithm. The mere idea of typicality implies a

real or virtual beholder (i.e., a human being or a simulating algorithm) that integrates the single exemplars encountered hitherto into a typical representation. For this reason, the notions of canonical or typical color are largely equivalent to the one of memory color.

Color Diagnosticity

While the term “memory color” characterizes a color, color diagnosticity characterizes an object ([12], for review, see [1, 2, 6]). It describes the degree to which an object is associated to, and hence characterized by, a particular color. We may distinguish objective and subjective color diagnosticity. In the objective sense, an object is color-diagnostic when it only occurs in a limited range of colors in the environment. For example, ripe Cavendish bananas only occur in a particular range of yellowish colors that is characteristic, or diagnostic, of ripe bananas. In contrast, cars are *color-neutral* (i.e., not color-diagnostic) because they occur in arbitrary colors. As the above example of the red Dacca banana showed, the objective co-occurrences of objects and colors and the subjective knowledge about them are not a priori the same. Subjective color diagnosticity describes the degree to which a beholder associates an object with a color, depending on the beholder’s prior experiences, familiarity, and criteria of typicality. Hence, subjective color diagnosticity implies that an observer has a memory color for an object.

Measurement

The method used to measure memory colors can make an important difference. For example, the word “lemon” is strongly associated to a yellow color. In contrast, an outline shape of a lemon might not be easy to recognize and is ambiguous because it could also represent a green lime instead of a yellow lemon (cf. [13]). The measurement of memory colors depends on three major characteristics: the identification of the object, the identification of the representative color (typicality), and the strength of the association between object and color (diagnosticity).

Object Identification

In order to measure a beholder’s memory colors for an object, the beholder must be able to recognize the object. This does not only require that the beholder is familiar with the respective kind of object, but it also depends on how well the exemplar used for the measurement may be identified as the respective kind of object. In particular, many measurements of memory colors use two-dimensional images, and the question arises whether the pictorial representations are recognizable and representative. For example, outline shapes of fruits, such as oranges or carrots, are barely recognizable even though people are highly familiar with these objects (cf. [13]). As a consequence, the measurement of memory colors may be affected by particularities of the image sampling. Some image databases address this problem by providing familiarity indices for the images, such as the Snodgrass line drawings [14].

Memory colors may also be investigated by describing objects verbally (i.e., by words or expressions). In this case, it is important to consider how specific and informative the term is about an object (e.g., fruit vs. banana vs. Cavendish banana), and how well observers know the meaning of those terms, e.g., the words “aubergine,” “eggplant,” and “*Solanum melongena*” refer to the same fruit, but may differ in comprehensibility.

Color Typicality

Memory colors consist of a most typical color that is associated with the respective object and a region of tolerance, within which concrete colors may still be accepted as instances of the memory color. The typical color may be represented by the average of all occurring colors, by the color that occurs in most cases (mode), by the color that is “purest” (i.e., maximally saturated), or simply by a whole region of colors (exemplar-based typicality).

In order to assess the range and typicality of colors associated with the objects, color adjustment or rating techniques may be used. In the former, observers adjust the color of the objects to the typical color, and the distribution of these adjustments informs about the nature and the

precision of the typical color (e.g., [3]). In rating tasks, observers evaluate how well a particular color matches the typical color of an object [15].

Color Diagnosticity

It is important to note that there may be an asymmetry in color-object associations: an object may refer to a specific color, even though the color might barely evoke the object. For example, while blue is highly diagnostic of a smurf [13], the first thing that comes to one's mind when thinking about blue is the sky or the sea rather than a smurf. The measurement of color diagnosticity and memory colors concerns the first case that is the strength, with which an object refers to a color.

A distinction may be made between conceptual and perceptual measures of color diagnosticity. A typical conceptual measure involves verbal elicitation. Observers are asked to name the most important object features, and color diagnosticity is assessed by the priority, with which color is named among these features. Other techniques require observers to rate how important color is to describe the object (reviewed, e.g., in [5]).

A perceptual approach consists in showing the respective objects (mostly their images) without color and asking the observer to indicate their typical color [13]. The accuracy and speed of the answer give insight into the strength of the object-color association. In particular, the speed may also reflect the automaticity of the memory color knowledge. Moreover, this technique also ascertains the validity of the measures for the particular stimuli. This, however, is also the drawback of this technique, since measurements cannot directly be generalized to other representations of the respective objects.

Characteristics of Memory Colors

Object and Scene Recognition

There is ample evidence that memory colors promote object and scene recognition. In particular, color-diagnostic objects and scenes are identified more easily when seen in their typical colors (for a review see [1, 5]). At the same time, chromatic

heterogeneity (cf. Fig. 1b) and three-dimensional shape cues increase the accuracy and precision of memory color measurements [11]. These observations suggest that memory colors and other visual characteristics of objects and scenes are tightly bound together and interact with each other.

Priming Effects

Furthermore, color-diagnostic objects automatically elicit the idea of the respective memory color and vice versa. In particular, so-called object-Stroop tasks have shown that colors are identified more rapidly when seen on the respective color-diagnostic object. In turn, priming paradigms revealed that colors facilitate the identification of the labels for color-diagnostic objects. Mentioning color-diagnostic objects verbally directs gaze automatically toward their memory colors. Similar to cueing, this implies an orientation of overt attention toward memory colors (for review, see [2]).

Color Constancy

It has been suggested that memory colors might help to establish color constancy, the ability to recognize surfaces across illuminations. The effect of the illumination color on the overall color of a scene might be estimated by comparing the actual color of a color-diagnostic object with its expected memory color. However, investigations of this idea only found very small effects of memory colors on color constancy, if any (for review, see [1, 2]).

Development

Interference effects that reveal memory color knowledge have been found at the age of 3.5 years with an object-Stroop-paradigm [9], and for 2- to 3-year-old children in a free-looking paradigm (e.g., [10]). A study using a preferential-looking procedure even found some evidence that infants possess memory colors as early as 6 months [8]. Nevertheless, preschool children (up to 5 years) may still have difficulties, when they are asked to identify the appropriate color for very common color-diagnostic objects (e.g., [7]).

The Memory Color Effect

According to the classical idea of a “memory color effect,” memory colors influence the appearance of the actual color of objects. The familiarity with the object-color association implies an expectation about the color of a concrete exemplar of an object, and this expectation contributes to the perception of its actual color. In this way, memory color effects indicate that color appearance depends on the object, on which the color is shown, as well as on the experience and the prior knowledge of the observer. For example, an observer may know that bananas are yellow. Does this knowledge influence how this observer sees the actual color of a banana?

Evidence

During the twentieth century, several studies pursued Hering’s original idea that memory colors affect the perception of the objects’ actual colors. Results of those studies were mixed, and it often remained ambiguous whether memory colors really affect perception rather than biasing memory retrieval and judgment (for review, see [2, 13]).

An achromatic adjustment approach was developed to assess genuine effects on color appearance [3]. If a color-diagnostic object automatically elicits the impression of its typical color, a banana should still appear to be yellow even if it is completely grey. To test this, participants were asked to adjust the colors of color-diagnostic

objects so that they look grey to them. If the object alone induces the impression of its typical color then the observers have to counteract this impression in order to see the object so that it looks subjectively grey to them. And indeed, the participants of those studies shifted the color adjustment toward the color opposite to the typical color. For example, they adjusted the banana slightly toward blue because blue is the color opponent to yellow. They did this, even though they had only to match the color of the object to the color of the grey background. This implies that they really saw the banana as yellow or at least slightly yellowish when it was actually grey. Figure 2 illustrates the idea and the rough size of the effect with a simple display taken from a follow-up study [16].

Determinants

Meanwhile, several studies have reproduced memory color effects using different methods and techniques. Besides achromatic adjustments, tasks included pairwise comparisons (as in Fig. 2), hue scaling, categorization, and the assessment of afterimages. Memory colors have been shown for natural objects, such as fruits and vegetables (e.g., [3]), for human-made objects, such as a Smurf and a German mailbox (e.g., [13]), and for scenes. At the same time, it also became clear that these memory color effects do not alter the ability to see color differences in discrimination tasks. Instead, they seem to affect how visible color information is subjectively experienced, maybe



Memory Color, Fig. 2 Memory color effect. The banana A is completely greyscale, while the banana B is slightly bluish. Most observers perceive A as yellowish and B as grey (from [16]). Note that the image is calibrated for an

sRGB monitor and may vary in color across display devices. On an sRGB monitor the difference between A and B approximates the memory color effect measured with the achromatic adjustment task [3]

similar to how attention modulates sensory information before it enters visual experience (for review, see [1, 2]).

Follow-up studies also allowed specifying factors that modulate memory color effects. Images with less perceptual information, such as outline shapes, yield weaker memory color effects. It has also been shown that memory color effects are strongest for objects with yellowish or bluish colors along the daylight locus, along which observers are least certain about their color appreciation. Red stimuli seem not to yield any memory color effects. These observations explain why some of the classical studies that used outline shapes and/or red objects yielded inconsistent findings (e.g., [13]; for review, see [1, 2]). How those determinants modulate the size memory color effects across stimuli might be understood through a Bayesian approach to the interaction between expectations and perception. A simple Bayesian model partially accounts for the variation of memory color effects [17]. Finally, a recent study found paradoxical memory color effects for faces under very particular viewing conditions: Faces look greenish when seen in the monochromatic yellow light of a (low pressure) sodium vapor lamp [18].

Neural Substrate

Neuropsychological studies found that brain injuries can affect object color knowledge (memory colors) independently from shape knowledge, color perception, and color naming, and that conceptual object color knowledge (“Bananas are yellow”) is independent from the ability to perceptually recognize whether an object is in its typical color (for review, see [4, 6, 23]). It has been suggested that perceptual information about memory colors are processed in an “anterior color center,” i.e., a parahippocampal, medial area of the cortex [4]. A recent single case study showed that a patient with color-specific deficits was able to verbally recall memory colors of objects and to recognize whether an object is shown in its typical color; but he was unable to judge whether a color patch corresponds to the memory color of an

object when the object is presented as a grayscale image separate from the color patch [19]. In-line with the aforementioned behavioral evidence, this observation suggests that memory colors are neurally represented as holistic templates (or “Gestalts”) that integrate perceptual color information with other visual information of the objects. Research into the neural basis of memory color effects indicates that greyscale images as well as verbal descriptions of color-diagnostic objects elicit color-specific activity in the primary visual cortex (V1) and the human visual area 4 (hV4) [20–22]. Taken together, these observations allow for the speculation that the perceptual memory color templates in the anterior color center modulate color experience through feedback connections to earlier visual areas. Nevertheless, the areas and their connections are still under research.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Fechner's Colors and Benham's Top](#)

References

1. Witzel, C., Gegenfurtner, K.R.: Color perception: objects, constancy, and categories. *Annu. Rev. Vis. Sci.* **4**(1), 475–499 (2018). <https://doi.org/10.1146/annurev-vision-091517-034231>
2. Witzel, C., Hansen, T.: Memory effects on color perception. In: Elliot, A.J., Franklin, A., Fairchild, M.D. (eds.) *Handbook of color psychology*. Cambridge handbooks in psychology, pp. 639–640. Cambridge University Press, Cambridge (2015)
3. Hansen, T., Olkkonen, M., Walter, S., Gegenfurtner, K.R.: Memory modulates color appearance. *Nat. Neurosci.* **9**(11), 1367–1368 (2006) nn1794 [pii], 10.1038/nn1794
4. Siuda-Krzywicka, K., Bartolomeo, P.: What cognitive neurology teaches us about our experience of color. *Neuroscientist.* (2019). <https://doi.org/10.1177/1073858419882621>
5. Bramao, I., Reis, A., Petersson, K.M., Faisca, L.: The role of color information on object recognition: a review and meta-analysis. *Acta Psychol.* **138**(1),

- 244–253 (2011). <https://doi.org/10.1016/j.actpsy.2011.06.010>. S0001-6918(11)00133-8 [pii]
6. Tanaka, J.W., Weiskopf, D., Williams, P.: The role of color in high-level vision. *Trends Cogn. Sci.* **5**(5), 211–215 (2001) S1364-6613(00)01626-0 [pii]
7. Gleason, T.R., Fiske, K.E., Chan, R.K.: The verbal nature of representations of the canonical colors of objects. *Cogn. Dev.* **19**(1), 1–14 (2004)
8. Kimura, A., Wada, Y., Yang, J., Otsuka, Y., Dan, I., Masuda, T., Kanazawa, S., Yamaguchi, M.K.: Infants' recognition of objects using canonical color. *J. Exp. Child Psychol.* **105**(3), 256–263 (2010). S0022-0965(09)00188-X [pii]. <https://doi.org/10.1016/j.jecp.2009.11.002>
9. Prevor, M.B., Diamond, A.: Color-object interference in young children: a Stroop effect in children 3(1/2)–6(1/2) years old. *Cogn. Dev.* **20**(2), 256–278 (2005). <https://doi.org/10.1016/j.cogdev.2005.04.001>
10. Johnson, E.K., McQueen, J.M., Huettig, F.: Toddlers' language-mediated visual search: they need not have the words for it. *Q. J. Exp. Psychol.* **64**(9), 1672–1682 (2011). <https://doi.org/10.1080/17470218.2011.594165>
11. Vurro, M., Ling, Y., Hurlbert, A.C.: Memory color of natural familiar objects: effects of surface texture and 3-D shape. *J. Vis.* **13**(7), 20 (2013). <https://doi.org/10.1167/13.7.20>
12. Biederman, I., Ju, G.: Surface versus edge-based determinants of visual recognition. *Cogn. Psychol.* **20**(1), 38–64 (1988) 0010-0285(88)90024-2 [pii]
13. Witzel, C., Valkova, H., Hansen, T., Gegenfurtner, K. R.: Object knowledge modulates colour appearance. *i-Perception.* **2**(1), 13–49 (2011). <https://doi.org/10.1068/i0396>
14. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn.* **6**(2), 174–215 (1980)
15. Smet, K.A.G., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**(3), 192–200 (2010)
16. Witzel, C.: An easy way to show memory color effects. *i-Perception.* **7**(5), 1–11 (2016). <https://doi.org/10.1177/2041669516663751>
17. Witzel, C., Olkkonen, M., Gegenfurtner, K.R.: A Bayesian model of the memory colour effect. *i-Perception.* **9**(3) (2018). <https://doi.org/10.1177/2041669518771715>
18. Hasantash, M., Lafer-Sousa, R., Afraz, A., Conway, B. R.: Paradoxical impact of memory on color appearance of faces. *Nat. Commun.* **10**(1), 3010 (2019). <https://doi.org/10.1038/s41467-019-10073-8>
19. Siuda-Krzywicka, K., Witzel, C., Taga, M., Delanoe, M., Cohen, L., Bartolomeo, P.: When colours split from objects: the disconnection of colour perception from colour language and colour knowledge. *Cogn. Neuropsychol.*, 1–15 (2019). <https://doi.org/10.1080/02643294.2019.1642861>
20. Bannert, M.M., Bartels, A.: Decoding the yellow of a gray banana. *Curr. Biol.* **23**(22), 2268–2272 (2013). <https://doi.org/10.1016/j.cub.2013.09.016>
21. Vandenbroucke, A.R., Fahrenfort, J.J., Meuwese, J.D., Scholte, H.S., Lamme, V.A.: Prior knowledge about objects determines neural color representation in human visual cortex. *Cereb. Cortex.* (2014). <https://doi.org/10.1093/cercor/bhu224>
22. Bannert, M.M., Bartels, A.: Human V4 activity patterns predict Behavioral performance in imagery of object color. *J. Neurosci.* **38**(15), 3657–3668 (2018). <https://doi.org/10.1523/JNEUROSCI.2307-17.2018>
23. Wang, X., Men, W., Gao, J., Caramazza, A., Bi, Y. Two Forms of Knowledge Representations in the Human Brain. *Neuron.* (2020). <https://doi.org/10.1016/j.neuron.2020.04.010>

Mesoamerican Color Survey Digital Archive

Kimberly A. Jameson¹, Nathan A. Benjamin²,
Stephanie M. Chang², Pruthi S. Deshpande³,
Sergio Gago⁴, Ian G. Harris⁵, Yang Jiao⁵ and Sean Tauber¹

¹Institute for Mathematical Behavioral Sciences,
University of California, Irvine, Irvine, CA, USA

²Calit2, Computer Science, University of
California, Irvine, Irvine, CA, USA

³Cognitive Sciences, University of California,
Irvine, CA, USA

⁴Calit2, School of Engineering, University of
California, Irvine, Irvine, CA, USA

⁵Computer Science, University of California,
Irvine, CA, USA

Definition

The Mesoamerican Color Survey (MCS) collected color-naming and categorization data from approximately 900 speakers from each of 116 indigenous languages from regions in Mesoamerica or Central America. Analyses of these data were originally reported by Dr. Robert E. MacLaury, principal investigator of the survey. The MCS data exist as a public-access color categorization and naming digital archive, in conjunction with other color categorization data

collected by MacLaury, and are available at <http://colcat.calit2.uci.edu/>.

Introduction

Human categorization behavior is widely studied across the behavioral sciences. It underlies many cognitive functions, including concept formation, decision making, learning, and communication. Color appearance, similar to other natural categorization domains, has distinctive features or properties that vary along continuous dimensions. Semantic color categories, their formation, their best exemplars and boundaries, and the influence of these on human behavior have been the topics of much empirical study, receiving considerable attention from anthropologist, linguists, cognitive scientists, and psychologists. The general aim of such research is to understand human color categorization and how it is cognitively and culturally represented across languages. For the case of color categories, the literature suggests that there are “universal trends” in human color representations with the category best exemplars being predictable across languages [1–7], and there are culturally specific color categorization influences [8–15], as well as human evidence [16–22] and simulated color category evolution evidence [23–27], suggesting combinations of universal, cultural, and pragmatic influences on color categorization and naming behaviors.

The Mesoamerican Color Survey (MCS) is one of the two existing databases (the other being the World Color Survey or WCS; see ► “World Color Survey”) which directly investigated, on a large scale, color naming and categorization across many linguistic societies. The MCS and the WCS employ nearly identical standardized procedures for evaluating large numbers of color stimuli, languages, and informants. The Mesoamerican Color Survey was conducted by Robert E. MacLaury during the years 1978–1981. MCS data were collected by MacLaury himself, or by research associates and colleagues in Mesoamerica, whom MacLaury trained and directed. The MCS data represent interviews with 900 speakers

of some 116 Mesoamerican languages. It is estimated that more than 100 indigenous languages are spoken in Mexico and Central America. Like most languages each MCS language has a color lexicon that partitions environmental color appearance stimuli according to a pattern that is specifically relevant to a given language’s speakers. But every local MCS system of color categorization also shares characteristics with systems observed for other Mesoamerican languages and with those of languages elsewhere in the world. MacLaury published analyses of the MCS data in the context of his *Vantage Theory* modeling approach ([28–30], see ► “Vantage Theory of Color”).

MCS and Theories of Color Categorization

Humans can discriminate on the order of 10^6 different colors, many more colors than any individual can name reliably. Most of the 7000 living languages have some form of a color lexicon. Various theories in the literature have sought to account for empirical findings on how each language represents color concepts and how such representations evolve, changing over time, and successively form hierarchical linguistic representations of color appearance similarity. In 1969 Brent Berlin and Paul Kay were among the first to systematically empirically examine how different linguistic societies represented color and theorize how color lexicons might similarly evolve ([1], see ► “Berlin and Kay Theory”). In the data they collected from 20 unrelated language families, they found (a) that there exist universal cross-linguistic constraints on color naming and (b) that basic color terminology systems tend to develop in a partially fixed order.

Berlin and Kay made two conjectures about the evolution of color lexicons: (1) There is a limited set of basic color terms (abbreviated as BCTs) in most languages, which are distinct from other color terms that an individual might use to name colors. (2) Color lexicons evolve from simple to complex, along highly constrained paths, starting

from two BCTs corresponding to warm-or-light and dark-or-cool categories in the most reduced lexicons and extending to 11 BCTs as seen in English and other languages spoken in industrialized societies. The Berlin and Kay theory is the basis for a mainstream approach to color categorization theory, along with its subsequent formulations using the World Color Survey (WCS) data ([2–4, 7], see ► [“World Color Survey”](#)).

While much empirical support exists for the Berlin and Kay/WCS theoretical approach, alternative explanations have also been suggested. Researchers directly involved in the WCS project, who collaborated with Berlin and Kay, have suggested that some linguistic societies might follow alternative patterns of color lexicalization, suggesting possible deviations from a hue-based model of color category universality that was central to the Berlin and Kay and WCS research program. The suggestions raised, for example, questioned whether a universal BCTs conjecture positing “a total universal inventory of exactly 11 basic color categories exists from which the 11 or fewer basic color terms of any language are always drawn” ([1], p. 2) was the best model to describe the extensive amount of data accumulated by the Berlin and Kay and WCS efforts. Robert E. MacLaury was among those who sought to explore alternative explanations, which was the impetus for his direction of the MCS project [11, 28–30]. In the 1970s MacLaury was an anthropologist working with Brent Berlin and Paul Kay, conducting fieldwork in Oaxaca, Mexico, and later working on the World Color Survey, and the Mesoamerican Color Survey was the basis for his PhD dissertation obtained from the University of California, Berkeley in 1986. In 1997 MacLaury successfully used the MCS data as the basis for an alternative color-naming and categorization theory, called Vantage Theory [28]. That approach presents a model of color categorization at the lexeme level where color categories are constructed as vantages (see ► [“Vantage Theory of Color”](#)). In the context of this approach, MacLaury’s book provides an extensive inventory of the organization and semantics of color categorization in Mesoamerica.

The MCS: History, Methodology, and Data

Historical Background

As noted earlier, the Mesoamerican Color Survey (MCS) includes 116 surveyed societies, each with a different Mesoamerican language or dialect, assessing approximately 900 native-language speakers from indigenous populations in Mesoamerica. The MCS employed color stimuli from the WCS stimulus palette shown elsewhere in this volume (see ► [“World Color Survey”](#)). Figure 1 above visually illustrates how the MCS substantially extends the geography of areas surveyed and thus extends beyond the wealth of information provided by the widely cited and valued WCS. It does so by replicating an estimated 20 languages found in the WCS, providing observations for 96 additional languages using the WCS’ standardized data collection procedures. Such an extensive survey of indigenous languages is likely no longer possible in 2015, due to dramatic increases in exposure over the last 25 years of native-language monolingual speakers to English-language broadcast media and entertainment.

The original MCS data was handwritten on paper, and while MacLaury published analysis and theory of MCS data (notably, [28]), thousands of pages of raw MCS data have never been publicly available and have not been subjected to the kinds of highly informative methodological approaches that have more recently been used to reveal the insights learned from the WCS data (e.g., [3–6, 20, 21, 31, 32]). MacLaury focused on MCS language groups across Northwest Mesoamerica for a particular reason: to investigate what he conceived as an alternative theory of color category systems that emphasized “brightness”-based categories in naming – as an alternative to historically prominent theories emphasizing hue-based categories in naming – and to test hypotheses and give interpretations of the data grounded in a relativistic “framing” idea, which MacLaury called Vantage Theory.

MacLaury employed the WCS methodology from 1978 up until approximately 2004. Subsequent to 2004 the archive containing the MCS was inactive, unlike that of the WCS, with its raw data

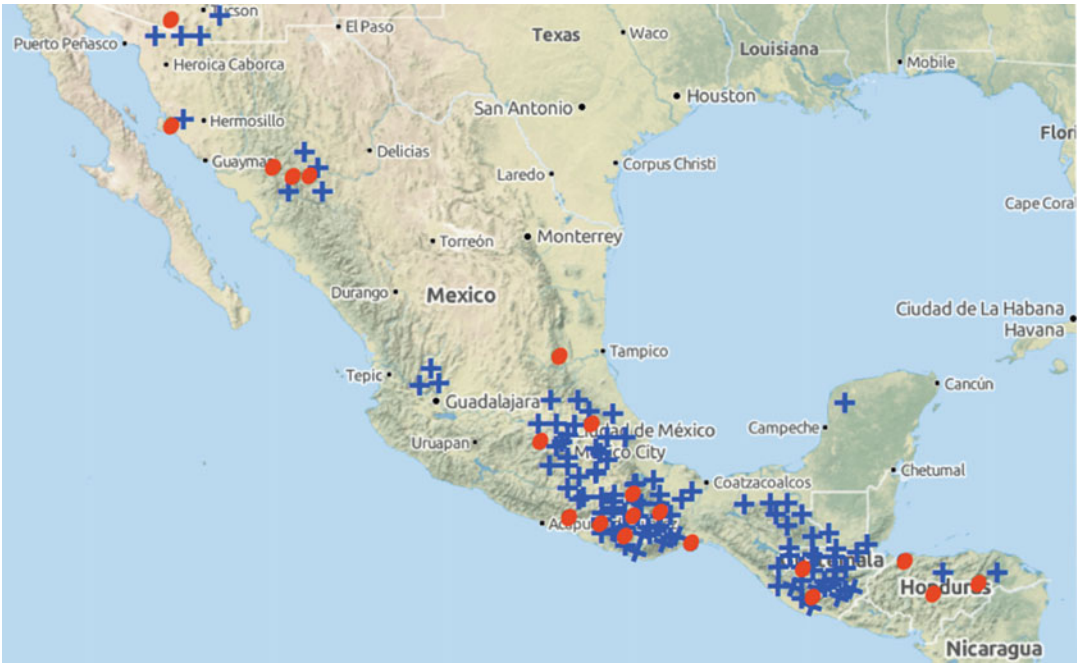
inaccessible to researchers while it was preserved in private storage by the MacLaury estate. Beginning in 2010 funding to initiate preservation and digitization of the archive was obtained. The initial preservation of the MCS archive, including its estimated 70 additional worldwide languages, came in the form of funding to purchase the archive from MacLaury’s estate, supported by the University of California Pacific Rim Research Program (award to K.A. Jameson, PI), and was augmented by a 2014 National Science Foundation funding (#SMA-1416907, K.A. Jameson, PI) that in part aimed to establish a public-access digital archive of transcribed MacLaury data and construct a user-friendly Wiki to generally facilitate collaborative color categorization research.

MCS Methodologies

Using individual samples and a color stimulus array shown in the WCS stimulus palette (see Fig. 1, Kay and Cook’s ECST entry), the MCS included three independent tasks done by every

native-language speaker surveyed. These consist of (1) a “naming task” that involved the free naming of 330 color samples presented in a fixed random order; (2) a “focus task,” identification of named category best exemplars or category “foci”; and (3) a “mapping task” that involved the demarcation of named category boundaries, producing a “boundary map.” The latter “mapping task” was used in Berlin and Kay’s original work, but for simplicity was omitted from the WCS. In the MCS, naming and focus data were collected for 900 individuals, and the mapping task method was reintroduced and assessed on an estimated 365 individuals. Data collection procedures used are detailed in works of MacLaury [28, 29] and are also similarly reviewed in WCS materials [2].

These MCS methods, along with the selection of languages it assessed, permit an unprecedented study of concept formation and linguistic representation, because of the following: (1) Many MCS languages are phylogenetically related to



Mesoamerican Color Survey Digital Archive, Fig. 1 Map depicting a portion of the geographic region surveyed by the MCS from the border of the United States to Nicaragua. Approximate locations of 116 MCS languages sampled are indicated as *blue crosses*, and 20 *red*

dots show WCS languages that duplicate some of the MCS languages surveyed. Of note are the dense samples of linguistically related MCS languages found in the areas of Oaxaca (37 languages), Guatemala (30 languages), and Mexico City (33 languages)

varying degrees. (2) Many MCS languages emphasize different perceptual dimensions (e.g., either primarily brightness or hue) and permit new analyses of dimensional salience. (3) Most MCS languages that satisfy (1) and (2) are related spatially, or geographically, and are thus known or expected candidates for linguistic borrowing, drift, and bilingual influences.

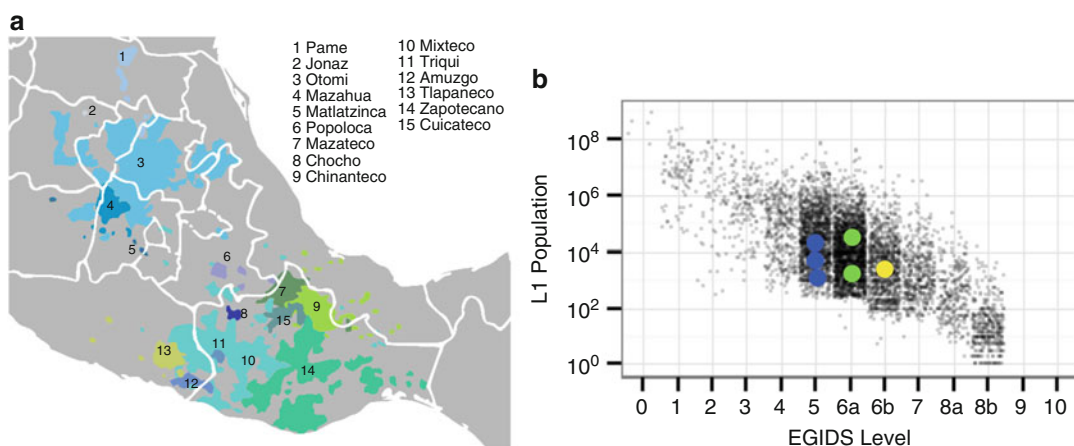
MCS Data

The research scope of the MCS data is substantial because it has features that can be exploited for addressing a range of heretofore unexplored color categorization research questions concerning how color categorization systems evolve, how semantic meaning drifts, and how meaning may be impacted by the influences of neighboring systems.

For example, Fig. 2's representation of one language family (Otomanguean, Glottolog: otom 1299) shows one example that is well represented in the MCS, and, like that shown in panel (a), the

MCS generally contains clusters of phylogenetically related languages, usually from contiguous geographic regions. Thus, for this one language family, as Fig. 2a illustrates, the MCS provides an estimated 369 surveyed participants, for which most subdivisions shown in panel (a) were assessed. Moreover, panel (b) illustrates that for some subdivisions, using Chinantec as an example, one may find several variations: That is, the MCS has six variations that cover three different levels of Chinantec language development/ endangerment.

In addition, the MCS also has many languages that emphasize different dimensions of color appearance space compared to the hue-based systems emphasized by languages typically investigated in the literature. For each language surveyed, the actual MCS datasets typically include detailed investigator notes or ethnographies, as well as the datasheets on the three tasks. Variations on information contained and



Mesoamerican Color Survey Digital Archive, Fig. 2 Example of the typical level of representation for one language family (Otomanguean) found in the MCS. Panel (a) shows a map of Oaxaca and Mexico City areas shown above in Fig. 1. Numbered color regions represent Otomanguean family language branches listed in inset (Retrieved on 12/02/2014 from “https://commons.wikimedia.org/w/index.php?title=File:Otomanguean_Languages.png&oldid=89424714”. Image: [wikipedia.org, https://commons.wikimedia.org/wiki/File:Otomanguean_Languages.png](https://commons.wikimedia.org/wiki/File:Otomanguean_Languages.png)) exemplifying spatial density of related languages that is typical of the MCS. Panel (b) shows one of the family branches found in (a), namely, #9 Chinantec. Panel (b)

shows a level of language diversity and complexity typical of the MCS. In this example, six forms of Chinantec data are illustrated in the MCS dataset. Panel (b)'s six color dots locate MCS Chinantec languages within the cloud of all living languages (i.e., *small dots*) in relation to a language's population (vertical axis) and its level of development or endangerment (horizontal axis). *Blue dots* are Developing (EGIDS 5) showing examples of Chinantec in vigorous use, *green dots* are Vigorous (EGIDS 6a) – unstandardized and in vigorous use among all generations. *Yellow dots* are In trouble (EGIDS 6b-7). (Chinantec living language data adapted and cited with permission from <http://www.ethnologue.com> [33]. Used by permission, © SIL International)

common formats of MCS datasheets are now described and illustrated:

1. **Naming task:** Informants were asked to name 330 loose color chips in a fixed random order. While investigators conducting the survey tended to employ two datasheet formats for collecting naming data, individual variations do exist in the raw datasheets. Figure 3 shows two common raw datasheet formats used to collect naming task data. Image scans of raw MCS datasheets, and for a currently limited number of cases their transcribed digital data files, are provided on the archive website (detailed below).

Figure 3a, b show two common examples of MCS naming task datasheet formats as raw pdf image scans of the data. Panel (a) illustrates only one page (samples 1–81) surveyed from informant #1 (assessed in Guarijio language) showing a “list format” in which the investigator sequentially listed a participant’s responses to 330 samples evaluated in the fixed random order in which they were assessed. Panel (b) shows a second common format in which the investigator reports the participant’s responses to 330 samples, evaluated in the same fixed random order, in the appropriate row/column location on the MCS stimulus palette (see Fig. 1 in ► “World Color Survey”). The palette format of the naming task data often has participant’s “focus” choices embedded in the datasheet as can be seen in panel (b)’s column 0 which has a “circled-X” symbol shown in cells *A0* and *J0*, recording the “foci” indicated by this participant for color appearances which in English language would gloss as the focus of “white” and the focus of “black,” respectively.

2. **Focus selection task:** Informants were shown the entire 330 color stimulus palette and asked to select the best example, or focus, of each different color name elicited in the naming task (see [28, 29] for method details). Again, individual variations do exist in the raw datasheets. Figure 4 shows two common raw datasheet formats used to collect focus selection task data. In panel (a) focus task data is reported in

conjunction with naming task data, whereas in panel (b) focus selection task data is combined with mapping task data (mapping task is detailed below). In both formats shown in Fig. 4, focus selections are indicated by either a “circled-X” symbol (panel (a)) or by a plain “X” symbol alone (as in panel (b)). In focus task datasheets, it is common to find a dictionary of color terms elicited at the bottom of the sheet (as seen in both Fig. 4 panels). Mapping task datasheets most frequently delineate mapped category areas using hatched markings (e.g., Fig. 4b); however these data were also often recorded using color pencils (as shown in Fig. 5 and described in the next section). As with the naming task data, image scans of raw MCS datasheets, and transcribed digital data files for focus data, are provided on the archive website.

3. **Category mapping task:** For the same color category terms as in the focus selection task, informants were shown the entire 330 color stimulus palette and asked to indicate the regions of the stimulus palette in which the appearance of a given color term was represented. The informants in this way would map the color category stimulus regions glossed by each color term for which an investigator solicited a mapping (see [28, 29] for method details). This task is also referred to as the rice mapping task in the literature, because grains of rice were often provided to informants to mark the many stimuli that were denoted by a given color term. Figure 5 shows two raw datasheets, for Guarijio and Chinantec languages, illustrating, above and beyond data variations, slight variations in the ways investigators recorded the surveyed information. Again, in some cases Focus locations are recorded, whereas in other Mapping Task datasheets Foci are not recorded. In mapping task datasheets, it is common to find a dictionary of color terms elicited at the bottom of the sheet and a pseudo-color code legend is given to interpret the regions defined using a particular color pencil as belonging to a specific color term (as seen in both Fig. 5 panels). (In such pseudo-color codes, the color of the

[illegible]

b

Name #1 Age 35 Sex M
Language Guarijío de Senora Baya Date 9/8/77
Locality La Junta, Alamos Stage IV
Senora, Mexico Investigator Stefan Mader

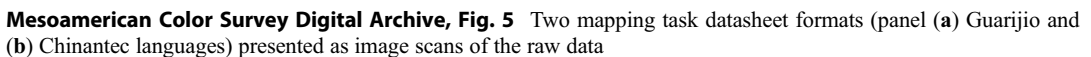
Mapping

	B	TR	Y	DY	G	H	SF	P
A	1	2	3	4	5	6	7	8
B	9	10	11	12	13	14	15	16
C	17	18	19	20	21	22	23	24
D	25	26	27	28	29	30	31	32
E	33	34	35	36	37	38	39	40
F	41	42	43	44	45	46	47	48
G	49	50	51	52	53	54	55	56
H	57	58	59	60	61	62	63	64
I	65	66	67	68	69	70	71	72
J	73	74	75	76	77	78	79	80

tokshname == siyóname ##### focus X
ohéoname ————— sa'wreme /////
sershname

Berlin and Kay Muscett Array
Color Category Code Sheet

Mesoamerican Color Survey Digital Archive, Fig. 4 MCS focus and mapping task datasheet formats shown as image scans of the raw data



ink used is not to be construed as resembling the color of the category data that it encodes.) As with the other MCS data, image scans of raw MCS datasheets are provided on the archive website.

The above details of MCS data, including Figs. 3, 4, and 5, convey the kind of coding variation found throughout the datasheets in the MCS archive. The variation is not wholly unexpected as a large number of the 116 languages of the MCS were collected by different investigators, at different locations, over the approximately 5 years during which the MCS was conducted. Notwithstanding this amount of idiosyncratic coding style, the MCS raw data, as well as the rest of the MacLaury archive, appear to be of rather consistent and thorough formatting concerning the data that is collected, especially considering the mass of information that exists in the estimated 20,000 pages of the archive.

The MCS and the Robert E. MacLaury Color Categorization Digital Archive (ColCat)

Recognizing the value in the MCS data, the interdisciplinary *ColCat* research group at UC Irvine sought to convert the paper copy of MacLaury's research archive into a public-access digital database, similar to that developed for the WCS by Paul Kay and colleagues [2]. The entirety of MacLaury's MCS data is included in this *ColCat* digital archive [34]. A major aim of the *ColCat* project is to develop a platform to make the archive, including the MCS data, readily available to the scientific research and teaching community. The project presently makes available organized sets of scanned image files of datasheet pages in the archive acquired from MacLaury's estate and continues to implement an intuitive Internet-based user interface to serve as the front end of a collaborative research space with which researchers can readily explore and investigate their own basic research questions concerning the MCS data, as well as the color

categorization data from other languages that MacLaury assessed. As mentioned earlier, neither the raw MCS data nor the MacLaury archive's additional non-Mesoamerican languages have been systematically organized for public use or previously published in an unanalyzed form. The additional 70 color categorization surveys (many with only a single informant, although others with frequently many more, and the largest with 40 informants) are valuable for their diversity in that they include native speakers from a wide range of languages including several Slavic languages, Hungarian, several Salishan languages of the Pacific Northwest United States, Zulu and several other South Africa/Zimbabwe languages, several native American languages, Germanic languages, European languages, Asian languages, and more.

The *Robert E. MacLaury Color Categorization (ColCat) Digital Archive* uses a content management system (CMS) for organizing, editing, and publishing the contents of the archive to a web-based graphical user interface (GUI) hosted at <http://colcat.calit2.uci.edu>. A partial list of features provided for the MCS through *ColCat* is summarized below:

1. User access level permissions and controls for the archive areas of the Wiki. User logins that define various levels of access for researchers, including those who may only want unregistered access to view database contents. Member level access for registered users who want to interact with the database. Editor level access as a trusted contributor to the database. Administrator access for systems operators who maintain database changes and review and approve contributions made by editor level users. Proper controls are implemented to maintain long-term integrity of the archive and to forestall loss and corruption of datafiles as well as accidental losses of data.
2. Facilitated search and query tools for exploring aspects of the datasets such as language type, language family, specific color terms, participant group subsets (e.g., by gender, age, etc.), survey

- geographic locations, and other filters that have been specified by content in the archived data.
3. Transcription quality indicators for evaluating conversion accuracy of handwritten data content into digitally addressable datasets. Because there are several ways the raw handwritten data in the archive can be converted into data addressable digital files (e.g., manual transcription, crowdsourced transcription, optical character recognition or OCR-based transcription, linguistic expert transcription, etc.), a useful feature of the ColCat Wiki includes utilities that permit assessment of each dataset's transcription quality and comparison across multiple forms of transcription of the same dataset. Transcription quality indicators appear on every surveyed language's archive page, and, where appropriate, indicate the degree of transcription completion, and computed degree accuracy, for each language's transcribed datasets.
 4. To encourage users to explore the archive's original raw datasheets, the *ColCat Wiki* provides each survey's scanned image files of the archive's raw data organized in a browsable catalog/library of pdfs (like those seen in Figs. 3, 4, and 5 above). Onboard pdf files of raw datasheets permit visual inspection of investigator's ethnographic notes and embellishments; verification of digitized data content; and querying pdfs by language family, language type, location and so forth, as well as querying by combinations of such tags.
 5. An onboard suite of supporting resources for color categorization research. For example, (a) on every surveyed language's archive page (where appropriate and provided for by the archive's contents), maps of the geographic regions are provided where surveys were conducted, permitting visualization of geographical relatedness among the archive's datasets, as well as allowing for search and identification of database contents by exploring geographical locations (via a Google Map plugin) where the actual survey data originated. (b.) Bibliographic resources highlighting specific languages or language families, including lists of peer-reviewed journal publications, books, relevant media, and reference materials, provided for each of the archive's languages surveyed. (c.) Active links out to relevant Internet resources, such as *World Atlas of Language Structures Database* (<http://wals.info>), and the public-access site *Ethnologue: Languages of the World Database* (<http://www.ethnologue.com>).
 6. As mentioned above a major aim of the archive project was to develop a platform to make the archive's data readily available to the scientific research and teaching community. Toward this end, the archive is designed to function as a collaborative research space, with user contributions enabled for trusted contributors, an informal blog feature to encourage a community of research exchange, and a platform for sharing and communicating results and for collective problem solving.
- Additionally important was the aim of making the digital archive's data accessible to a broad audience comprised of users with varying levels of expertise in handling and analyzing large amounts of data. For this reason, a set of *ColCat data tools* was developed which allow users to search, filter, and analyze data in ways that best meet their research goals and technical abilities.
- The planned data-handling tools and utilities aim to facilitate, for all user levels, the use of the archive's contents for original color categorization research studies.
- The ColCat digital archive, including the MCS, contains data at three levels:
1. Digital image scans, in pdf format, of approximately 12,500 handwritten pages of color-naming data, ethnographic notes, and support documents from 116 Mesoamerican languages, plus pdf image scans of an additional 10,000 pages surveyed globally from other languages and nations
 2. Raw transcription data of these surveys converted and collected through OCR, crowd sourcing, and expert transcribers, and

3. “Official” transcription conversions for the archive’s color-naming and categorization data

The Wiki’s onboard data tools are designed for visualizing and analyzing the computer-addressable archive data at levels 2 and 3. Visualization tools for transcription data (level 2) include quantitative indicators of completeness and quality of transcriptions for each language and visualizations of the variance between or within transcription methods for entire languages or subsets of images within a language. Visualization tools for color-naming data (level 3) will include visualizations of the aggregated color-name mappings for each language including information about the proportion of agreement among all respondents or subsets of respondents through the use of filters – i.e., based on gender, age, etc.

Data export tools will allow users to download both transcription and naming data (levels 2 and 3) in formats conducive to off-line analysis (i.e., CSV, Excel, etc.). Although information about aggregated color-name mappings (level 3) will be based on data from all subjects in a language, users may be interested in computing and/or comparing consensus color-name mappings for subpopulations based on their own criteria. To that end, downloadable tools will allow users to take the same analyses and visualizations that are used on the Wiki for complete populations and apply them to subpopulations that were exported by the user. The expectation is that a variety of user types will employ the Wiki’s visualization tools to help identify subsets of data in the archive that meet specific research goals, while other researchers will be more inclined to export data in a familiar format (i.e., in an established WCS format) and apply their own off-line analyses. Regardless of the case, the ColCat data tools make the entire archive’s data accessible to a wide range of researchers, students, and other scientists who would like to explore and learn from the archive but are not well equipped to manage and process large quantities of data.

Uses of the MCS Archive

As a complementary database to the World Color Survey database, the addition of the MCS archive allows for an even greater, more extensive, assessment of behaviors of cross-cultural color cognition and color naming. For example, reconsider the Fig. 2a discussion above, concerning the 369 surveyed Otomanguean language family informants for which MCS data exists. These MCS informants can now be independently compared to, or even analyzed in conjunction with, the 175 previously surveyed WCS informants from the Otomanguean language family. In addition to the novel surveys it adds, it is this kind of duplication and augmentation that the MCS provides as a complementary database to the WCS that will increase the value of both archives and present new opportunities to greatly extend understanding of how the data in the MCS and the WCS inform on the questions originally asked by the surveys’ directors, namely, what specifically are the universal and culturally specific patterns of color naming and categorization that can be empirically demonstrated for large samples of the world’s diverse languages and how do they inform us on the larger question of general human tendencies for categorizing and naming objects perceived in the world.

Cross-References

- ▶ [Berlin and Kay Theory](#)
- ▶ [Bilingual Color Naming/Categories](#)
- ▶ [Color Categorical Perception](#)
- ▶ [Color Category Learning in Naming-game Simulations](#)
- ▶ [Color Dictionaries and Corpora](#)
- ▶ [Dynamics of Color Category Formation and Boundaries](#)
- ▶ [Effects of Color Terms on Color Perception and Cognition](#)
- ▶ [Vantage Theory of Color](#)

References

- Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
- Kay, P., Berlin, B., Maffi, L., Merrifield, W.: *The World Color Survey*. Center for the Study of Language and Information, Stanford (2003)
- Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
- Regier, T., Kay, P., Cook, R.: Focal colors are universal after all. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8386–8391 (2005)
- Lindsey, D.T., Brown, A.M.: Universality of color names. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16609–16613 (2006)
- Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within-language diversity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19785–19790 (2009)
- Kay, P., Regier, T.: Color naming universals: the case of Berlinmo. *Cognition*. **102**, 289–298 (2007)
- Davidoff, J., Davies, I.R.L., Roberson, D.: Color categories in a stone-age tribe. *Nature*. **398**, 203–204 (1999)
- Roberson, D., Davies, I.R.L., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
- Roberson, D., Hanley, J.R.: Color categories vary with language after all. *Curr. Biol.* **17**, 605–606 (2007)
- MacLaury, R.E.: Color-category evolution and shuswap yellow-with-green. *Am. Anthropol.* **89**(1), 107–124 (1987)
- Paramei, G.V.: Singing the Russian blues: an argument for culturally basic color terms. *Cross-Cult. Res.* **39**(1), 10–38 (2005)
- Dedrick, D.: Color language universality and evolution: on the explanation for basic color terms. *Philos. Psychol.* **9**(4), 497–524 (1996)
- Jameson, K.A.: Culture and cognition: what is universal about the representation of color experience? *J. Cogn. Cult.* **5**(3–4), 293–347 (2005)
- Alvarado, N., Jameson, K.A.: Confidence judgments and color category best exemplar salience. *Cross-Cult. Res.* **39**(2), 134–158 (2005)
- Jameson, K.A.: Why GRUE? An interpoint-distance model analysis of composite color categories. *Cross-Cult. Res.* **39**(2), 159–194 (2005)
- Jameson, K.A.: Where in the world color survey is the support for the hering primaries as the basis for color categorization? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 179–202. The MIT Press, Cambridge (2010)
- Davies, I.R.L., Corbett, G.G.: A cross-cultural study of color grouping: evidence for weak linguistic relativity. *Br. J. Psychol.* **88**(3), 493–517 (1997)
- Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. *PLoS One*. **8**(2), e55986 (2013). <https://doi.org/10.1371/journal.pone.0055986>
- Bimler, D.: Are color categories innate or internalized? Hypotheses and implications. *J. Cogn. Cult.* **5**(3), 265–292 (2005)
- Bimler, D.: From color naming to a language space: an analysis of data from the world color survey. *J. Cogn. Cult.* **7**(3), 173–199 (2007)
- Bimler, D., Uusküla, M.: Clothed in triple blues: sorting out the Italian blues. *J. Opt. Soc. Am. A.* **31**, A332–A340 (2014)
- Narens, L., Jameson, K.A., Komarova, N.L., Tauber, S.: Language, categorization, and convention. *Adv. Complex Syst.* **15**(3–4), 1150022 (2012)
- Jameson, K.A., Komarova, N.L.: Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am. A.* **26**(6), 1414–1423 (2009). Featured Reprint in *The Virtual J. Biomed. Opt.* **4**(8)
- Jameson, K.A., Komarova, N.L.: Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A.* **26**(6), 1424–1436 (2009). Featured Reprint in *The Virtual J. Biomed. Opt.* **4**(8)
- Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *J. Theor. Biol.* **253**, 680–700 (2008)
- Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)
- MacLaury, R.E.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, Austin (1997)
- MacLaury, R.E.: *Color in mesoamerica*. Vol. 1: a theory of composite categorization. Doctoral dissertation. University of California, Berkeley. UMI University Microfilms, No. 8718073, Ann Arbor (1986)
- MacLaury, R.E.: From brightness to hue: an explanatory model of color category evolution. *Curr. Anthropol.* **33**(2), 137–186 (1992)
- Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language*. **85**, 884–892 (2009)
- Webster, M., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R., Paramei, G., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
- Paul, L.M., Simons, G.F., Fennig, C.D. (eds.): *Ethnologue: Languages of the World*, Eighteenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> (2015)
- Jameson, K.A., Gago, S., Deshpande, P.S., Benjamin, N.A., Chang, S.M., Tauber, S., Jiao, Y., Harris, I.G., Xiang, Z., Bhakta, H.R., MacLaury, R.E.: The Robert E. MacLaury Color Categorization (ColCat) Digital Archive. <http://colcat.calit2.uci.edu/>. The California Institute for Telecommunications and Information Technology (*Calit2*), UC Irvine (2015)

Metal Halide Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Definition

A lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a high-pressure mercury vapor with metal halide additives that is contained in a transparent bulb.

Metal Halide Gas Discharge Lamps

Metal halide lamps are high-pressure mercury gas discharge lamps that contain metal halides in addition to the mercury [2, 3]. In the heated discharge tube, the metals of the halides take part in the discharge process and radiate their own spectrum. Compared with high-pressure mercury lamps, both color properties and efficacy are considerably improved. Thanks to the fact that no fluorescent powder is needed, the small gas discharge tube itself is the light-emitting surface. This small light-emitting surface makes the lamps extremely suitable for use in reflector and floodlight luminaires. Initially, they were solely produced in extremely high lumen packages with relatively large lamp dimensions (about the size of a 1-liter bottle). These lamps are particularly suited for use in floodlights for the lighting of stadiums. Today, compact metal halide lamps are also available in small lumen packages, which make them very suitable for use in compact reflector luminaires for accent lighting both indoors and outdoors. These versions have become an energy-efficient alternative for halogen lamps. Special compact metal halide versions are being produced for use

in film studios and theatres. The smallest metal halide lamps are used for car headlamps (these are the so-called xenon lamps, whose gas discharge tube is no larger than a match head). Metal halide lamps belong to the group of high-intensity discharge (HID) lamps because they are available in high-lumen output (and thus high-luminous-intensity) versions.

Working Principle

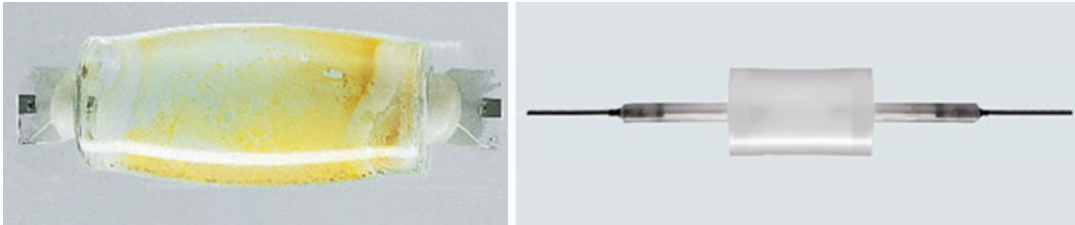
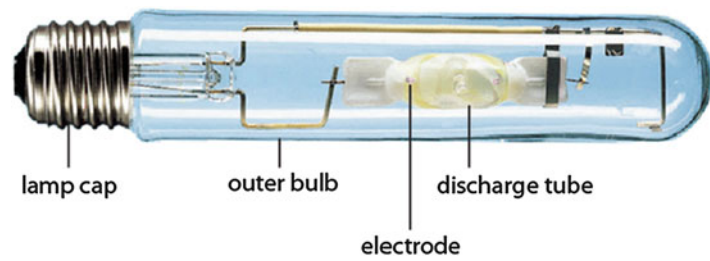
Suitable metals that vaporize in the hot discharge tube so as to contribute to the discharge process cannot be added directly to the mercury in the discharge tube. This is because metals suitable for this purpose attack the wall of the gas discharge tube at the high temperatures that are needed for the gas discharge to occur. The solution to this problem is to add these metals in the form of their nonaggressive chemical compounds with a halogen (iodine, bromide, or chloride), hence the name metal halide lamps. The solid metal halide starts evaporating once, after switching on the lamp, the mercury discharge has increased the temperature in the discharge tube. When this metal halide vapor enters the area of the mercury discharge in the center of the tube with its very high temperature (around 3,000 °C), it dissociates into its separate elements: metals and halogen. There the metals in their pure, vaporized state take part in the discharge process and determine the efficacy and color characteristics of the radiation. The aggressive vaporized metal cannot reach the tube wall because, at the lower wall temperature (some 1,000 °C), they recombine again to form the harmless metal halide compound. Like all gas discharge lamps (with very few exceptions), a metal halide lamp cannot be operated without a ballast to limit the current flowing through it. It also needs, again like most gas discharge lamps, an igniter for starting the lamp.

Materials and Construction

The main parts of a metal halide lamp are (Fig. 1):

- Discharge tube
- Metal halide additives
- Fill gas

Metal Halide Lamp,
Fig. 1 Main parts of a
 metal halide lamp [1]



Metal Halide Lamp, Fig. 2 Gas discharge tube made of quartz (*left*) and ceramic material (*right*), respectively. The yellow material seen through the quartz tube is condensed metal halide [1]

- Outer bulb
- Electrodes
- Lamp cap

Discharge Tube

The discharge tube is made of either quartz or ceramic material. Some metals of the metal halide compounds (especially sodium) have the tendency, at high temperatures, to migrate slowly through the quartz wall of the tube, with the result that there is a gradual change in the color properties of the lamp during its lifetime. The use of ceramic discharge tubes solves the problem. Such tubes are impervious to these metals, even at high operating temperatures. High-pressure sodium lamps make use of the same ceramic material. Metal halide lamps using ceramic material are referred to as ceramic metal halide discharge lamps. Figure 2 shows examples of both quartz and ceramic gas discharge tubes. Since the ceramic tube is produced by sintering minuscule aluminum-oxide particles, one cannot look straight through the wall, although the light transmittance is more than 90% (translucent instead of transparent material). Ceramic gas discharge tubes are operated at a higher temperature than are quartz tubes. The higher operating

temperature influences the spectrum of the radiation. The result is a lower color temperature compared to that of quartz metal halide lamps and a 10% higher efficacy.

Metal Halide Additives

In theory, some 50 different metals can be used for metal halide compounds, and different manufacturers have introduced various combinations of these metals. Examples of some of the metals used in metal halide compounds are sodium, thulium, thallium, indium, scandium, dysprosium, and tin. Most lamps use a mixture of at least three different metal halides. Each different combination results in a different spectrum, but also in different efficacy, lumen maintenance, and lifetime characteristics.

Fill Gas

Besides metal halides and mercury, an inert gas is also added to the discharge tube, like in most other discharge lamps.

Electrodes

The electrodes of metal halide lamps are of the type used in ordinary high-pressure mercury lamps. They consist of a tungsten rod, with a

tungsten coil impregnated with emissive material, wound around it. They are of more massive construction than the ordinary high-pressure mercury lamps because of the higher operating temperature.

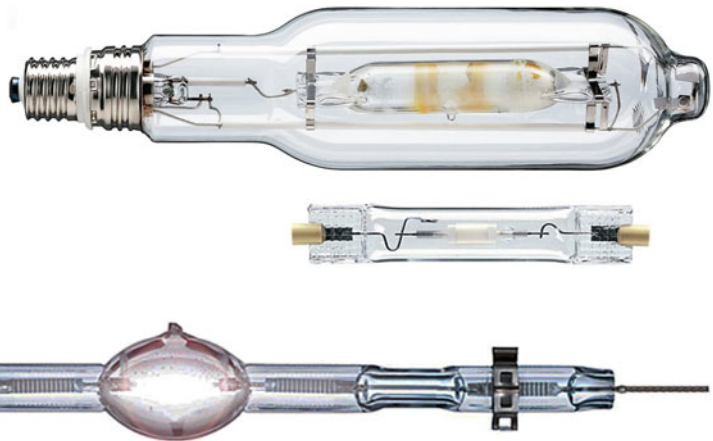
Outer Bulb

Most metal halide lamps use a hard glass or (for very compact versions) quartz outer bulb for protection and for heat insulation of the discharge tube. Both single-ended and double-ended outer bulbs are used (Fig. 3). They may be evacuated or gas-filled. In the case of quartz outer bulbs, UV-blocking quartz is used to limit harmful UV radiation. For the same reason, some hard glass versions use a cylindrically shaped UV-block shield around the discharge tube.

The inner wall of the outer bulb reflects a very small part of the light that consequently cannot be very well controlled by a luminaire reflector. But since the reflected amount is so small, this usually is not a problem, except where very-high-lumen-output lamps are employed. In floodlight installations, a very small amount of uncontrollable light could give rise to disturbing light pollution. For these applications special lamps are therefore available that have no outer bulb (Fig. 4). For reasons of thermal stability and safety, those lamps without an outer bulb must be used in specially designed luminaire housings.

Metal Halide Lamp,

Fig. 3 *Top* tubular hard glass, single-ended outer bulb and *bottom* quartz, tubular double-ended outer bulb



Metal Halide Lamp, Fig. 4 Double-ended quartz metal halide lamp not making use of an outer bulb

Lamp Caps

Metal halide lamps come with a great variety of lamp caps (Fig. 5). The single-ended lamps of higher wattage have in general E40 Edison screw caps. Some caps have special electrical insulation because of the high ignition pulses of between 600 and 5,000 V. The lower-wattage single-ended lamps have two-pin electrical connections of various types, sometimes with ceramic electrical insulation. In order to be able to accurately position the gas discharge arc in a luminaire, some lamps are provided with a so-called prefocused lamp cap (Fig. 5, right). The double-ended lamps use lamp caps of the type that are also used in double-ended halogen lamps (Fig. 3, bottom). The lamps without an outer bulb again have different lamp caps (Fig. 4).

Properties

Energy Balance

Almost 25% of the input power of metal halide lamps is emitted in the form of visible radiation. Compare this with 15–17% of normal high-pressure mercury lamps and 30% of high-pressure sodium lamps.

System Luminous Efficacy

The compact versions of the metal halide lamp have lumen efficacies (depending on the metal

Metal Halide Lamp,**Fig. 5** Various lamp caps for compact single-ended metal halide lamps

halide mixture and gas discharge tube material used) of between 70 and 95 lm/W. When comparing this with ordinary halogen incandescent lamps with their maximum efficacy of 25 lm/W, it is clear that compact metal halide lamps are often very suitable to replace halogen lamps. The larger versions of metal halide lamps have efficacies from some 75 lm/W to slightly more than 105 lm/W.

Lumen Package Range

Compact versions are available in lumen packages from some 1,500 to 25,000 lm (corresponding wattage range 20–250 W). Larger versions range from some 20,000 lm to more than 200,000 lm (corresponding wattage range 250–2,000 W). Special (compact) types for use in film studios, theatres, and professional photography are produced in lumen packages of up to more than 1,000,000 lm (12 kW).

Color Characteristics

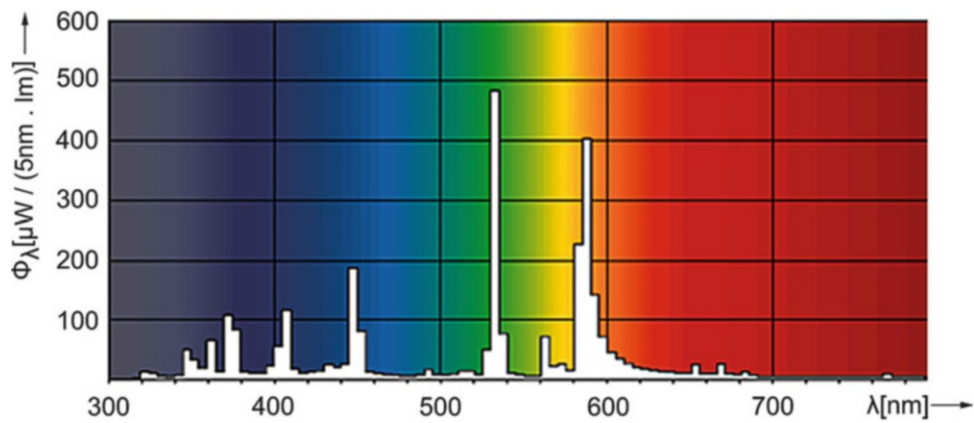
As has been explained in the sections above, by employing different metal halide mixtures, lamps with different spectra can be produced. Like all gas discharge lamps, the metal halide lamp has a discontinuous spectrum. Figures 6, 7, and 8 show the spectra of some typical metal halide lamps with different color temperatures T_k and color rendering index R_a . The same color

designation system is used as with fluorescent lamps.

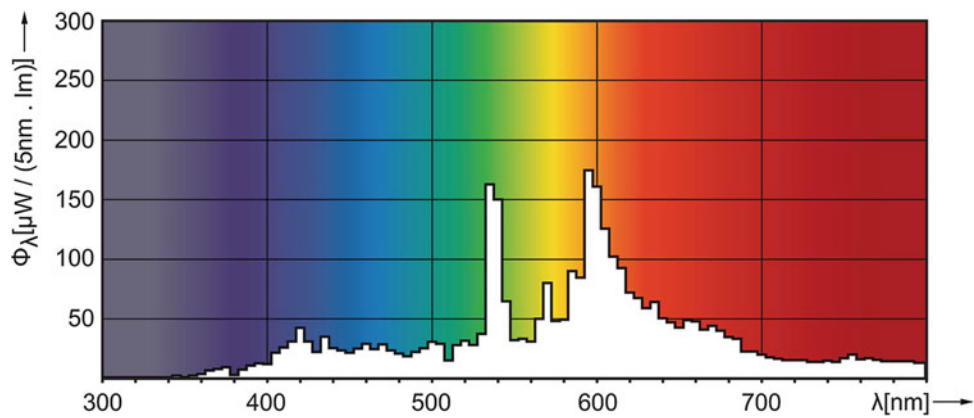
Compact metal halide lamps are typically produced in the color temperature range of 3,000 K to some 4,500 K in two color rendering varieties: R_a ca. 80 and ca. 90. The larger metal halide lamps are available in the color temperature range from approximately 4,000 to 6,000 K, with color rendering R_a in the range from 65 to more than 90. The special, daylight, metal halide lamp types for film studio and theatre lighting have high color temperatures of between 6,000 and 8,000 K with color rendering values R_a between 65 and more than 90.

Lamp Life

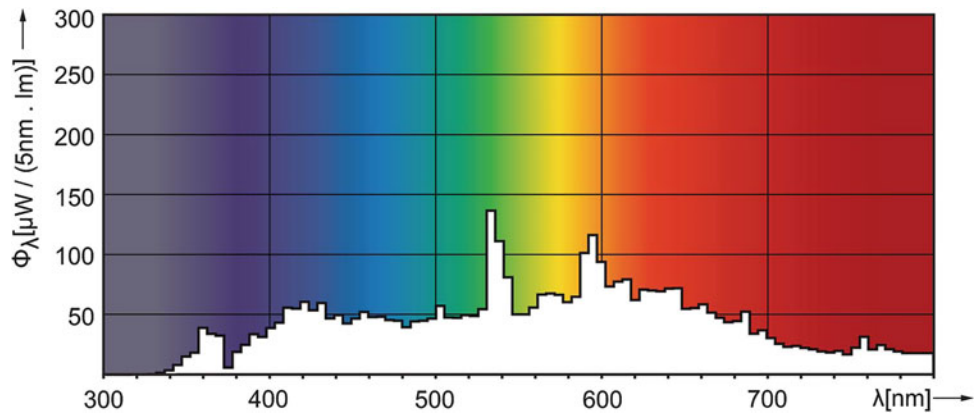
The lamp life of most metal halide types is somewhat shorter than that of other gas discharge lamps. This is because the electrodes are heated to a higher temperature with a correspondingly higher evaporation rate and are gradually worn out by chemical reactions with the metal halides. It has also been shown that different metal halide lamp types employ widely different construction methods. As a consequence, lamp life varies strongly with type. The economic life of compact versions lies between some 7,000 and 14,000 h (20% mortality). The high-lumen-package versions vary from 4,000 h (single-envelope types for stadium floodlighting) to some 10,000 h (again



Metal Halide Lamp, Fig. 6 Spectral energy distribution of a metal halide lamp, color type 642: T_k 4,200 K and R_a 65 [1]



Metal Halide Lamp, Fig. 7 Spectral energy distribution of a ceramic metal halide lamp, color 830: T_k 3,000 K and R_a 80 [1]



Metal Halide Lamp, Fig. 8 Spectral energy distribution of a ceramic metal halide lamp, color 942: T_k 4,200 K and R_a 90 [1]

20% mortality). These ceramic metal halide lamps specifically developed for use in road lighting, where lights may be used 4,000 h a year, have lifetimes up to 20,000 h (20% mortality).

Lamp Lumen Depreciation

Metal halide lamps have a higher lumen depreciation than most other discharge lamps. This is because of the higher degree of blackening from evaporated electrode material. Here too, the type of construction and sort of metal halides used to play a role. Lumen depreciation values vary between some 20% and 30% (after 10,000 h). Some special types, such as those employed in road lighting, depreciate less rapidly (some 10% after 10,000 h).

Burning Position

With the larger gas discharge tubes, in particular, the burning position may affect the actual location of the various metals in the tube. This means that different burning positions may result in different color shifts. Also, with some types of lamp construction, the burning position can influence the lifetime of the lamp, for example, because of attack of the electrodes by some of the metals of the halides. Many metal halide lamps, therefore, have restrictions as to their permitted burning position (these are, of course, specified in their accompanying documentation). Compact, single-ended types, with either quartz or ceramic tube, have universal burning positions.

Run-Up and Reignition

The metal halides in the discharge tube need time to heat up, evaporate, and dissociate into metal and halide. During this process, which takes about 2–3 min, the light output and color gradually change until the final stable condition is reached. If there is an interruption in the power supply, medium- and high-wattage lamps will take approximately 10–20 min for the pressure in the lamp to decrease enough for it to reignite. Compact ceramic lamps reignite much faster: after some 3–5 min. Immediate reignition is only possible by applying a very high-voltage pulse

(typically 60 kV). For this purpose, special, hot restrike, lamps are produced with special heavily insulated electrical contacts for applying this high-voltage pulse.

Dimming

The dimming of metal halide lamps is difficult because, with the resulting decrease in temperature, some of the metal halides condense, changing the constitution of the metal halides actually participating in the discharge. This, in turn, changes the color properties of the light. By employing a specially shaped burner, electronically driven lamp versions have been developed that can be dimmed to some 50% without suffering from this problem.

Mains Voltage Variations

Mains voltage variations affect lumen output, lifetime, and light color. A mains voltage that deviates 10% from its nominal value will result in perceivable color shifts. Thanks to the higher operating temperature of ceramic lamps, the effect on color change here is considerably smaller. Both up and down variations in the mains shorten lamp life.

Metal Halide Lamp, Table 1 Different metal halide lamp types

Lamp parameter	Range
Lamp size	Medium/compact/very compact
Discharge-tube material	Quartz/ceramic
Lamp shape	Single-ended/double-ended
Outer bulb	Outer bulb/coated outer bulb/no outer bulb
Lumen package	Large/small
Lamp circuit	Electromagnetic/HF electronic
Color temperature T_k	3,000–4,000 K/4,000–7,000 K
Color rendering index R_a	$R_a > 90$ (900 series)/ $80 < R_a < 90$ (800 series)/ $R_a < 80$
Lifetime	Standard (7,000–14,000 h)/long (up to 20,000 h)
Reflector	No reflector/built-in reflector



Metal Halide Lamp, Fig. 9 Examples of metal halide lamps used mainly in indoor lighting



Metal Halide Lamp, Fig. 10 Examples of metal halide lamps used mainly in outdoor lighting. *First three:* for road and industrial lighting. *Last four:* for sports floodlighting

Product Range

Metal halide lamps are available in a wide variety of types for many different applications, including floodlighting, road lighting, accent lighting (both interior and exterior), car head lighting, and film studio lighting. The more important lamp parameters, together with the corresponding range for which different versions are made, are listed in Table 1. Figure 9 shows some examples of lamp types used mainly in indoor lighting, while Fig. 10 shows those types primarily used in outdoor lighting.

Cross-References

► [High Pressure Mercury Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Coursebook. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. Illuminating Engineering Society of North America, New York (2011)

Metamerism

Peter van der Burgt

Philips Lighting, Valkenswaard, The Netherlands

Definition

Metamerism is the matching of apparent color of objects with lights that have different *spectral power distributions*. Colors that match are called metamers.

Color of the Light

The light coming from any light source (sunlight, incandescent lamp) can be described by the so-called spectral power distribution, often denoted as SPD. The SPD gives the amount of power radiated by the light source as a function of wavelength in the visible range of the electromagnetic spectrum being 380–740 nm. Every SPD will generate a color impression in the human visual system. To describe the color of the light, a system has been developed by the CIE (Commission Internationale de l'Eclairage). A set of three color-matching functions called $x(\lambda)$, $y(\lambda)$, and $z(\lambda)$ has been defined, and these are known collectively as the CIE standard observer [1]. These color-matching functions are depicted below in Fig. 1.

By using the color-matching functions, the effect of a *spectral power distribution* $E(\lambda)$ can be expressed by the following equations:

$$X = \sum E(\lambda) \cdot x(\lambda) \quad (1)$$

$$Y = \sum E(\lambda) \cdot y(\lambda) \quad (2)$$

$$Z = \sum E(\lambda) \cdot z(\lambda) \quad (3)$$

The values X , Y , Z are called the tristimulus values of the color of the light with spectral power distribution $E(\lambda)$.

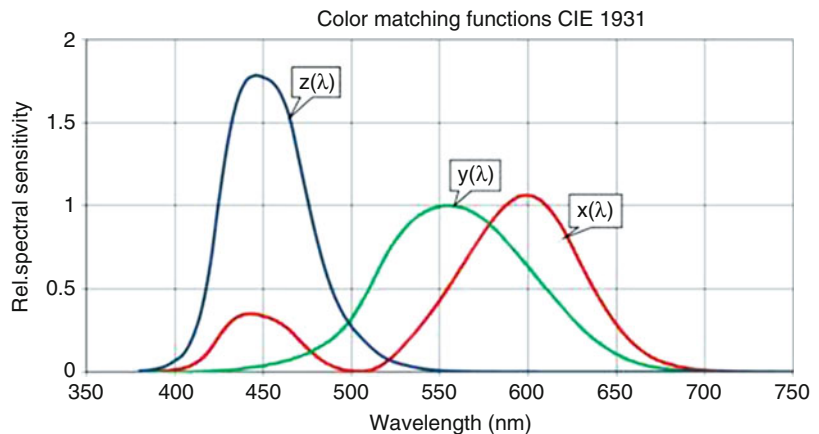
The tristimulus values for a light can be converted to the chromaticity or color point x , y by the following equations:

$$x = X/P \text{ with } P = X + Y + Z \quad (4)$$

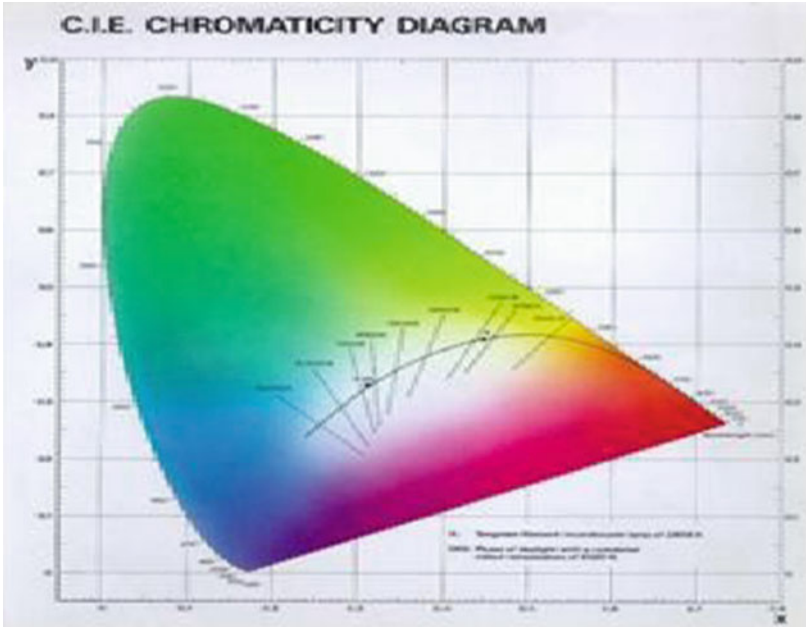
$$y = Y/P \quad (5)$$

These equations lead to the well-known CIE chromaticity diagram which is shown in Fig. 2. With the equations above, the color points for monochromatic lights can be calculated. This leads to the spectral locus, the outer curved boundary in the diagram. The straight edge on the lower part of the gamut is called the line of purples. These colors, although they are on the border of the gamut, have no counterpart in monochromatic light. Less saturated light colors appear in the interior of the figure. The diagram shows in

Metamerism,
Fig. 1 Color-matching
functions CIE



Metamerism, Fig. 2 CIE chromaticity diagram



the colored area all the chromaticities that are visible to the average person, and this colored region is called the gamut of human vision. One important property of this diagram is that lights that have the same chromaticity will give the same color impression to an observer. The chromaticity, however, does not predict the precise color impression of a light. For example, an incandescent lamp burning in the middle of the day may give a yellowish impression. The same lamp in the night will yield a white impression. So the precise color impression will depend on the adaptation of the observer. But at the same adaptation, the same chromaticity will lead to the same color impression.

Metamerism of Light Colors

Equations 1, 2, and 3 show that it is possible that more lights having different SPDs can nevertheless have the same values for the tristimulus values X, Y, Z and, therefore, the same chromaticity. Such lights are called metamer; they have different spectral power distributions but give the same color impression. It should be noted that this

metameric character applies within the boundaries of the CIE XYZ system.

Metamerism of Object Colors

The spectral composition of a color stimulus of an object color can be described by the following equation:

$$E_R = \beta(\lambda) * E(\lambda) \tag{6}$$

with

$E(\lambda)$: spectral power distribution of the light that illuminates

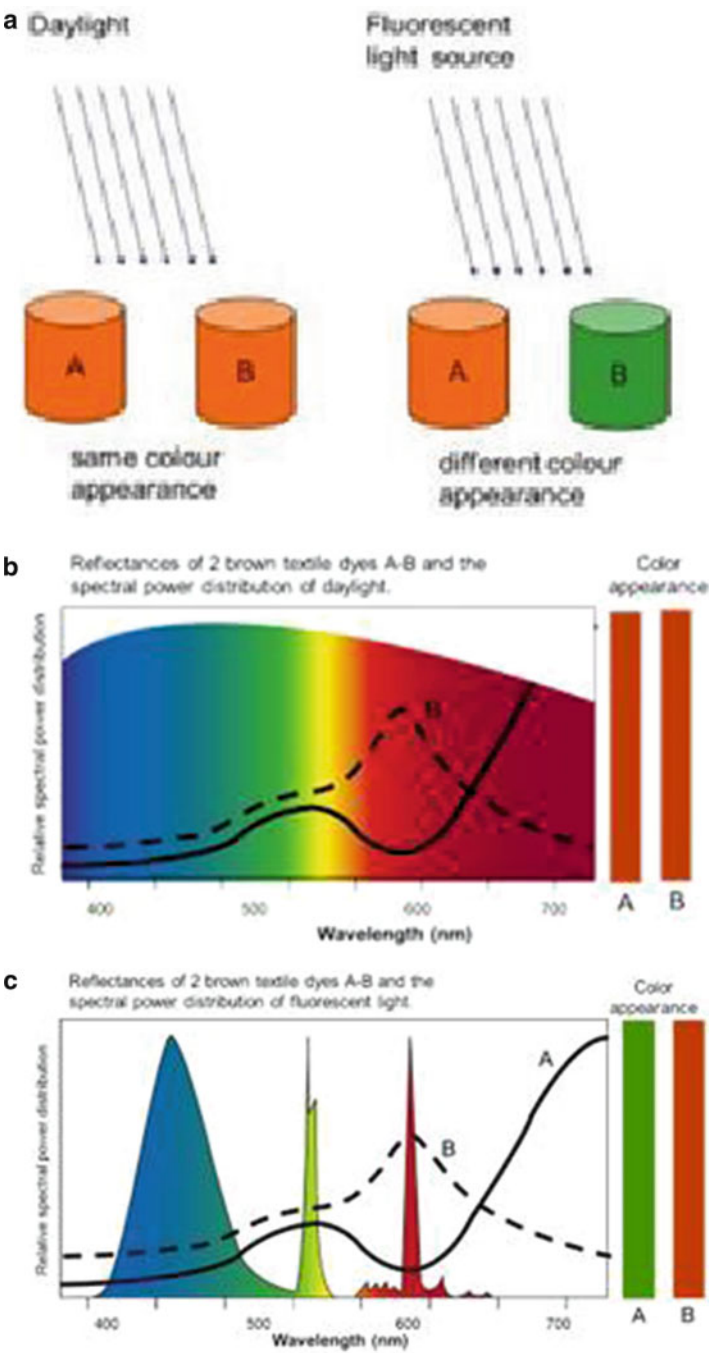
$\beta(\lambda)$: spectral reflectance of the object (fluorescent objects will be neglected here)

The tristimulus values of this color stimulus are as follows:

$$X = \sum \beta(\lambda) * E(\lambda) . x(\lambda) \tag{7}$$

$$Y = \sum \beta(\lambda) * E(\lambda) . y(\lambda) \tag{8}$$

Metamerism, Fig. 3
(a–c) Change of light source leads to change in color appearance



$$Z = \sum \beta(\lambda) * E(\lambda).z(\lambda) \tag{9}$$

Suppose the amount of light on the object is decreased with a factor 10, so $E(\lambda) = 0.1E(\lambda)$. From experience, it is known that the human

visual system adapts in such a way that the color impressions stay the same.

Now suppose that the spectral reflectance is reduced with a factor 10, so $\beta(\lambda) = 0.1 \beta(\lambda)$. Again from experience, it is known that the color

impression will change: a white will turn into a gray, an orange color will turn brownish, etc.

How to account for these two observations? This can be done by defining the tristimulus values of object colors by the following equations:

$$X = 100 * \sum \beta(\lambda) * E(\lambda) . x(\lambda) / \sum E(\lambda) . y(\lambda) \quad (10)$$

$$Y = 100 * \sum \beta(\lambda) * E(\lambda) . y(\lambda) / \sum E(\lambda) . y(\lambda) \quad (11)$$

$$Z = 100 * \sum \beta(\lambda) * E(\lambda) . z(\lambda) / \sum E(\lambda) . y(\lambda) \quad (12)$$

By defining the values like this, they have become invariant for multiplication of $E(\lambda)$ with a constant factor, so invariant for the light level.

At the same time, multiplication of $\beta(\lambda)$ with a constant factor will lead to accordingly changes in the tristimulus values of the object color.

Another way of doing this is by normalizing $E(\lambda)$ in such a way that

$$\sum E(\lambda) . y(\lambda) = 100 \quad (13)$$

Normalizing $E(\lambda)$ like this enables the use of Eqs. 7, 8, and 9 to calculate the tristimulus values.

For light sources, the light color can be defined with only two coordinates, the chromaticity x, y .

For color stimuli of object colors, however, three coordinates are necessary. These can be X, Y, Z or x, y, Y or any other three-dimensional system.

Object colors will be metameric if the three tristimulus values of the colors will be equal.

Suppose two object colors have spectral reflectances $\beta_1(\lambda)$ and $\beta_2(\lambda)$. These object colors will be metameric under a light $E(\lambda)$, normalized like in Eq. 13, if

$$\begin{aligned} X_1 &= \sum \beta_1(\lambda) * E(\lambda) . x(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) . x(\lambda) = X_2 \end{aligned}$$

$$\begin{aligned} Y_1 &= \sum \beta_1(\lambda) * E(\lambda) . x(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) . x(\lambda) = Y_2 \end{aligned}$$

$$\begin{aligned} Z_1 &= \sum \beta_1(\lambda) * E(\lambda) . x(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) . x(\lambda) = Z_2 \end{aligned}$$

These object colors are metameric under the light with spectral power distribution $E(\lambda)$. It is clear that the metameric character depends on the light source. In general, with another light with a different SPD, the relations $X_1 = X_2, Y_1 = Y_2$, and $Z_1 = Z_2$ will no longer hold. This means that colors metameric under one light source, that is, giving the same color impression, can look rather different under a light source with a different SPD. This is illustrated in Fig. 3. Two brown textiles with different reflectances will have the same color appearance in daylight, whereas in a specific fluorescent light, the color appearance will be different. In Fig. 3b, c, the reflectances of the colors are depicted together with the spectral power distributions of daylight and fluorescent light.

In the paint industry, metameric object colors are very well known. In this industry, a frequently occurring problem is to reproduce a color with other pigments than the ones used for the original color. In such cases, usually samples are sought that are metameric under daylight as well as incandescent light.

Cross-References

- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Color Rendering Index](#)
- [CIE Special Metamerism Index: Change in Observer](#)

References

1. ISO 11664-1:2007(E)/CIE S 014-1/E:2006: Joint ISO/-CIE Standard: Colorimetry Part 1. CIE Standard Colorimetric Observers

Mie Solution

► [Rayleigh and Mie Scattering](#)

Mie Theory

► [Rayleigh and Mie Scattering](#)

Minimally Distinct Border

Michael H. Brill
Datacolor, Lawrenceville, NJ, USA

Definition

In an exactly abutting pair of spatially uniform colored regions, if one of the regions is varied in radiance until the border between the regions is least conspicuous to a human observer, that border becomes a minimally distinct border (MDB) and is seen as having maximal blur and minimal contrast. The term MDB is used both for the border itself and for the psychophysical method used to derive it.

Historical Beginnings

The minimally distinct border was first reported by Boynton and Kaiser [1] in connection with psychophysical experiments. From the beginning, it was realized that the condition of MDB occurs when the two abutting regions have the same luminance. Hence, the condition of MDB is best understood in terms of the properties of luminance. (See next section.)

However, the history of the MDB is not just about luminance [2, 3]. The MDB was first used in an attempt to quantify complete color differences regardless of their size. Previously, whereas small color differences could be quantified by

discrimination thresholds, large color differences were assessed by magnitude estimation. Compared to magnitude estimation, the assessment of strength of a visual border seemed to offer greater precision. In such experiments, subjects were asked to define the achromatic contrast that they deemed as equivalent to the chromatic contrast of a given border. When the radiances were titrated so as to produce a condition of MDB – hence equality of luminance – one could then look at the strength of the border as a function of the other color coordinates. Historically, MDB experiments tended to compare a monochromatic light with a reference white, so it was thought that the MDB in such cases quantified saturation or departure from the white. As technology developed to present more light at shorter wavelengths (470 nm and lower), it was observed that the MDB could quantify only a “tritanopic saturation,” because no short-wave-sensitive (blue-sensitive) photoreceptors contributed to MDB. In fact, the MDB was observed to vanish completely for lights that were the same in all but their short-wavelength content – hence the name “tritanopic,” which derives from the nomenclature of color blindness. To quote Boynton [2]: “The blue-sensitive cones [...] seem to be free of any serious spatial or temporal responsibilities in vision.”

Luminance and MDB

Luminance of a light was initially thought to be linear in the light’s radiance and to transform perceptually to its brightness. Following this definition, two spectrally different lights were declared to have the same luminance if the radiance of one light was adjusted so its brightness matched the brightness of the other light. Early in the twentieth century, a more precise definition for luminance was devised: Two lights are equally luminous if, when the lights are presented alternately, the perceived flicker vanishes at the lowest frequency of alternation. The practice of finding equally luminous lights using this definition is called flicker photometry.

To be in harmony with both these definitions, in 1924, the Commission Internationale de

l'Eclairage (CIE) developed a hybrid definition based on experiments of both sorts: flicker fusion of two alternating lights and direct brightness comparison of two lights at neighboring wavelengths [4]. As defined by the CIE, the luminance of any light is proportional to the integral of its spectral power distribution weighted by a function $V(\lambda)$ (called the luminosity function) derived from these experiments.

In the next 30 years, it was discovered that $V(\lambda)$ could not capture all the facts about isoluminous lights. For one thing, Deane Judd found in 1951 that $V(\lambda)$ should be modified to have a larger contribution at short wavelengths. Also, the physical size of the light stimulus (field size) was found to have an influence on luminosity, as testified by the 1964 CIE colorimetric standard that was tailored to a 10° field size rather than to the 2° field size previously used. Finally, the luminosity of lights at low (scotopic) radiance follows the spectral sensitivity of the rods rather than the cones of the visual system. The low-light limit of luminance also entailed a transition to the high-light limit through a radiance domain known as the mesopic, in which both rods and cones operate.

Clearly the 1924 CIE definition of luminosity, although ubiquitous in industrial and other practical usage, diverges from what is found in actual visual systems. Therefore, vision researchers prefer to define luminance according to an empirical criterion, such as implied by flicker photometry. Once this definition is adopted, it is found that various other visual phenomena depend only on luminance of the stimulating light. In particular, visual acuity is a function only of luminance. In fact, the evidence supports the characterization of luminance as a channel in human vision, fed by long- and medium-wavelength-sensitive cones, occasionally by rods, and never by the short-wavelength-sensitive cones. The luminance channel is tuned to stimuli of the highest spatial and temporal frequencies perceivable by the visual system.

Using the vision-science definition of luminance described above, the condition of MDB between two abutting light patches occurs when the patches have the same luminance.

Measuring MDB

An MDB experiment finds the relative luminance of a light field as follows. For a fixed reference field, the MDB radiance of a monochromatic test field at wavelength λ is measured. The ratio of the reference to test radiance at MDB is proportional to the test field's luminance as a function of wavelength.

The MDB used in this way can be an empirical definition of luminance. In fact, because MDB presents the test and reference stimulus simultaneously, it has been touted as superior to flicker photometry as a method of measuring luminance.

MDB studies have been done over a variety of geometric conditions. Typically, the experiments involve a bipartite field of two semicircles (test field and reference field). The semicircles meet at a diameter on which the subject is directed to fixate. Alternatively, if one wants to compare the strength of an MDB with the strength of an achromatic border, one can divide the field into quarters of a circle, the top two comprising the achromatic border and the bottom two comprising the chromatic MDB.

It should be noted that creating the geometric conditions for MDB is not trivial. For example, an achromatizing lens is recommended to remove chromatic aberration, which would otherwise disturb the edge between the abutting fields.

Physiological Basis

The measurement of MDB was possible in macaque retina under several experimental conditions, most notably when the border was moved back and forth in the macaque's visual field. Two kinds of retinal mechanisms, fast (phasic) and slow (tonic), were examined as possible vehicles for MDB, and evidence favored the phasic mechanism [5].

As with any psychophysical measure of luminance, additivity (i.e., Abney's law) has been found to be imperfect, but this is hardly surprising since the luminance channel in vision receives inputs from more than one cone channel, and cone response is nonlinear in light intensity.

Furthermore, the low-luminance limit must engage the rods as well as the cones, so additivity cannot be a strict property of MDB.

Extension to Mesopic Photometry

Raphael and MacLeod [6] discovered that equating luminance at various mesopic radiance levels (in which both rods and cones enter the function) is correlated strongly with achieving MDB at these levels.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [Chromatic Contrast Sensitivity](#)
- [Color Categorization: Patterns and Mechanisms of Evolution](#)
- [Luminance Meter](#)

References

1. Boynton, R.M., Kaiser, P.K.: Vision: the additivity law made to work for heterochromatic photometry with bipartite fields. *Science*. **161**, 366–368 (1968)
2. Boynton, R.M.: Ten years of research with the minimally distinct border. In: Armington, J.C., Krauskopf, J., Wooten, B.R. (eds.) *Visual Psychophysics and Physiology*. Academic, New York (1978)
3. Boynton, R.M.: *Human Color Vision*. Holt, Rinehart, Winston, New York (1979)
4. Wyszecki, G., Stiles, W.S.: *Color Science*, 2nd edn. Wiley, New York (1982)
5. Kaiser, P.K., Lee, B.B., Martin, P.R., Valberg, A.: The physiological basis of the minimally distinct border demonstrated in the retinal ganglion cells in the macaque retina. *J. Physiol.* **422**, 153–183 (1990)
6. Raphael, S., MacLeod, D.I.A.: Minimally distinct borders in mesopic vision. *J. Vis.* **8**(17), article 75 (2008)

Mixed Adaptation Condition

- [CIE Guidelines for Mixed Mode Illumination: Summary and Related Work](#)

Mixed Chromatic Adaptation

- [CIE Guidelines for Mixed Mode Illumination: Summary and Related Work](#)

Modern Art and Fashion Design

- [Art and Fashion Color Design](#)

Monochromator

Yi-Chen Chuang

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Definition

A monochromator is an optical dispersing device that is used to select a narrow band of light (i.e., optical radiation) from a wider range of wavelengths available at the input. The Greek roots “mono-” and “chroma-” refer to “single” and “color” respectively. Ideally, a monochromator should produce a single wavelength of optical radiation at its output. Although lasers produce light that is much more monochromatic than the optical monochromators discussed here, only some lasers are easily tunable but are not as simple to use.

Monochromators are included in many optical measurement instruments and systems for applications where tunable monochromatic light is required. A monochromator combined with optical detectors can be used to obtain the spectral power distribution (SPD) of light sources, reflectance or transmittance of objects, etc.

Monochromators are important for color measurement because many color-related optical characteristics are dependent on wavelength.

In color science, reflected or transmitted light from a sample is usually measured, where a monochromator can be used before or after the incident light illuminates a sample. If a monochromator is used before the incident light illuminates the sample, the incident light is monochromatic light. On the other hand, if a monochromator is used after the light illuminates the sample, the monochromator is used to convert the reflected or transmitted light to monochromatic signals for analysis. It is typically used in a spectrometer (or spectroradiometer) or a spectrophotometer.

There are different types of monochromator based on its color selection mechanisms and/or designs, e.g., prism, Czerny–Turner, holographic grating, double, etc. The principles of these monochromators are introduced in the “[Overview](#)” section below.

Overview

The main function of a monochromator is to separate the color components of a light. It can use either the optical dispersion phenomenon in a prism or that in a diffraction grating.

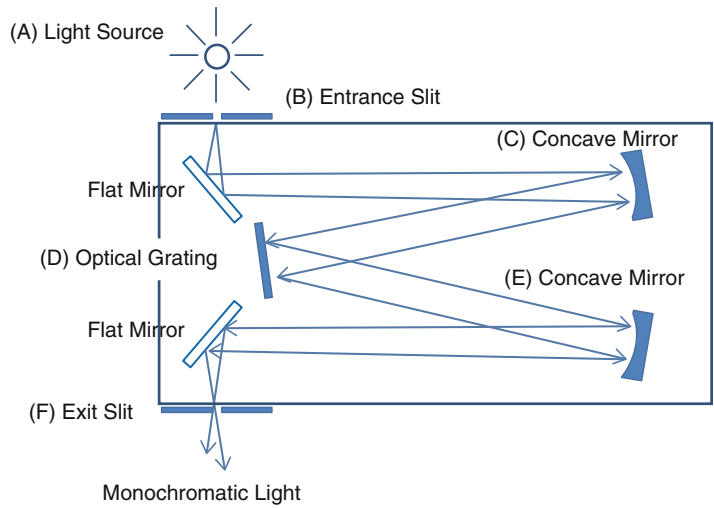
Figure 1 shows a scheme of a simple monochromator. Various types of monochromator have been developed, but a monochromator usually contains an entrance slit, an essential dispersing element, and a mechanism to direct the selected color to an exit slit for color selection purposes [1].

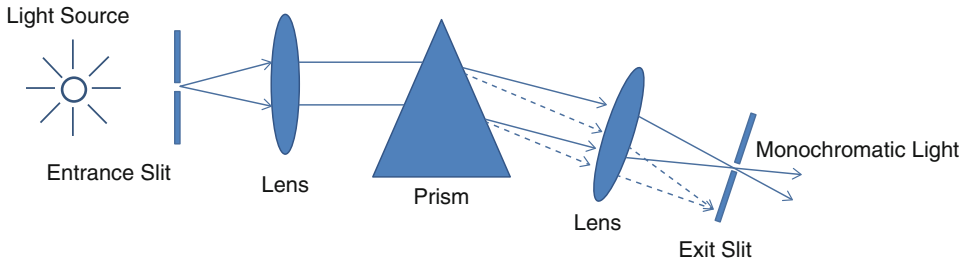
The Czerny–Turner monochromator is a type of monochromator using diffractive gratings as the dispersing element. A Czerny–Turner monochromator is sometimes referred to as a single monochromator because the radiant flux is only diffracted once [2]. A schematic diagram of a



Monochromator, Fig. 1 Scheme of a simple monochromator

Monochromator, Fig. 2 Basic structure of a Czerny–Turner monochromator





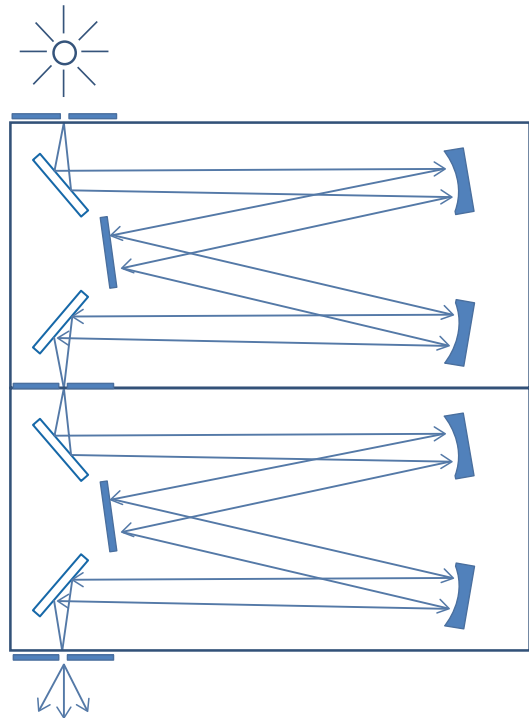
Monochromator, Fig. 3 An example of a prism monochromator

Czerny–Turner monochromator is shown in Fig. 2. In Fig. 2, the light under measurement (A) is incident through the entrance slit (B), reflected by a concave mirror (C) to become collimated. It then illuminates the optical grating (D), disperses light to different colors, reflected and focused by the concave mirror (E) to the exit slit (F). By rotating the optical grating (D), different colors of light can reach and exit at the exit slit (F), and such colors are “selected” by the monochromator. The orientation of the optical grating (D) and the spatial location of the exit slit (F) determine which wavelengths of the light are selected. The two concave mirrors (C and E) image the entrance slit on the exit plane. In practice, the collimating and focusing elements (C and E) are usually the same mirror, used twice [2].

The double monochromator usually contains two Czerny–Turner monochromators used in a system. Double monochromators are known for their double dispersion and thus can significantly reduce scattered light (Fig. 4) [2].

A prism monochromator uses a prism as the dispersing element (Fig. 3). It is not necessary for the collimating and focusing elements to be concave mirrors, but they can be optical lens. Usually, reflective mirrors are preferred over optical lens because mirrors conserve more optical radiation energy during the transmission process. However, because of the higher cost and small angular dispersion, prism monochromators are rarely made today (Fig. 4) [2].

A holographic grating monochromator uses holographically constructed concave gratings to simplify the optics. This eliminates some of the lenses and mirrors that would normally be



Monochromator, Fig. 4 Double monochromator

required with conventional flat plane grating monochromators.

When choosing a monochromator, the following specifications should be considered: wavelength range, wavelength accuracy, measurement speeds, spectral bandwidth, spectral scattering, slit widths, grating groove density (if applicable), mechanical and thermal stability, size and weight, stray light, and dispersion [3, 4]. For example, Table 1 shows the criteria for the selection of gratings according to the above specifications. The resolution is the minimum detectable difference

Monochromator, Table 1 The criteria for the selection of a grating [4]

Specification of monochromators	Criterion
<i>Groove density (or groove frequency):</i> The number of grooves contained on a grating surface (lines/mm)	Groove density affects the mechanical scanning range and the dispersion properties of a system. It is an important factor in determining the resolution capabilities of a monochromator. A higher groove density results in greater dispersion and a higher resolution Select a grating that delivers the required dispersion, such as that for a charge-coupled device (CCD), array detector, or a required resolution (with an appropriate slit width) when using a monochromator
<i>Mechanical scanning range:</i> The wavelength region in which an instrument can operate	Refers to the mechanical rotation capability (not the operating or optimal range) of a grating drive system with a specific grating installed Select a grating groove density that allows operation over the desired wavelength region
<i>Blaze wavelength:</i> The angle in which the grooves are formed with regard to the normal grating, often called the blaze angle	Diffraction grating efficiency plays an important role in spectrograph throughput. Efficiency at a particular wavelength is largely a function of the blaze wavelength if the grating is ruled, or a modulation if the grating is holographic Select a blaze wavelength that encompasses the total wavelength region of the applications. More attention should be paid to covering the short wavelength of the spectrum region
<i>Quantum wavelength range:</i> The wavelength region of highest efficiency for a particular grating	Normally determined by the blaze wavelength Select a grating with maximum efficiency over the required wavelength region for the applications

between two spectral peaks provided by a monochromator. Theoretically, resolution is approximately equal to slit width (mm) $\times$ dispersion (nm/mm), where the dispersion describes how well the light spectrum is spread over the focal plane of the exit slit. However, in practice, because

of aberrations in monochromators, the actual resolution is rarely equal to this. The actual resolution of a monochromator as a function of the slit width is described as the bandwidth (nm) of a monochromator.

When using a monochromator, some parameters need to be set up properly, as they affect the measurement results. For example, smaller slit widths produce smaller spectral bandwidths, which corresponds to a narrower spectral band at the exit slit (i.e., closer to the ideal single-wavelength output). However, smaller slit widths mean that less optical radiation power is transmitted through the monochromator, which results in lower output signals and decreases the signal-to-noise ratio of the signals measured. The bandwidth of commonly used monochromators is from the order of 10^{-2} to 10 nm. Detailed information on monochromators can be found in manufacturers' user manuals.

It is important to evaluate the performance of a monochromator to reduce measurement uncertainties (i.e., gain more confidence regarding the accuracy of the results measured). To evaluate the performance of a monochromator, the following characteristics should be considered: wavelength accuracy, spectral bandwidth, stray light, system reproducibility, and system linearity.

Cross-References

- [Instrument: Spectrophotometer](#)
- [Spectral Power Distribution](#)
- [Spectrometer](#)
- [Spectroradiometer](#)

References

1. Webster, J.G. (ed.): The Measurement, Instrumentation, and Sensors Handbook. Springer, Heidelberg (1999)
2. Christensen, R.L., Potter, R.J.: Double monochromator systems. *Appl. Optics*. **2**, 1049 (1963)
3. Kostkowski, H.J.: Reliable Spectroradiometry. Spectroradiometry Consulting, La Plata (1997)
4. Yoshizawa, T. (ed.): Handbook of Optical Metrology: Principles and Applications. CRC Press, Boca Raton (2009)

Mosaics

Gertrud Olsson

Academy of Design and Crafts at Steneby,
University of Gothenburg, Gothenburg, Sweden

Synonyms

[Tessellations](#); [Tesserae](#)

Definition

Arrangement of small fragments of stones, colored glasses, golden pieces, or other materials, fixed on walls or pavements and depicting an image or making an abstract pattern. The technique of mosaic is a decorative art that reached a peak during the Byzantine period of Christian art.

Overview

The traditional mosaics consist of marbles and stones in the scale of black and white (sometimes together with red) or other materials which were available in nature. At the beginning, in the Antique age, mosaics were used only as pavements, as stones on the ground, a floor material. During the first centuries of Christianity, during the Byzantine period, the Church took over the mosaic technique. The Byzantine technique with walls done with mosaics was new in the fifth and sixth centuries. In the basilicas of Ravenna, Italy, the Byzantine mosaic masters worked with entire walls of mosaic made up of small pieces of colored glass and gold, *tesserae*, affixed to mortar (Fig. 1). *Tesserae*, a term derived from a Greek word meaning “four-sided,” is the standard for mosaic pieces. Before being divided, the tesserae are part of a blown flat glass in the form of a “pizza plate” (in today’s terminology). *Smalti* is the technical term for the brilliant, opaque-colored crystalline material fused with glass.

In Turkey not least in Istanbul, we find the glass mosaic tradition from Antiquity and



Mosaics, Fig. 1 San Apollinare Nuovo (526), Ravenna. (Photograph by the author)

Byzantine times. Marble mosaic as a flooring material from the excavated parts of the late antique Great Palace is now in the Mozaik Müzesi in Sultanahmet (Fig. 2). The Byzantine style, with large amounts of glass mosaics, can be seen in the Hagia Sophia (ninth to thirteenth centuries) (Fig. 3), Chora/Kariye Museum and Pammakaristos Church/Fethiye Museum (both fourteenth century), decorated entirely on walls and ceilings, and in apses and domes.

The new materials that the Byzantine mosaic masters introduced were colored glass and gold. These materials offer a wider and more intense color scale than the marble mosaics. The tesserae-mosaic is a kind of color material that is mixed optically (on the retina of the eye). The material interacts in a special way with vision and light. The surface is shiny, hard, and reflective. Glass pieces are shiny and slightly irregular in pure colors and gold. The gold picks up light and reflects gold light back into the room. The entire composition involved the setting of mosaic pieces at different



Mosaics, Fig. 2 Great Palace (fifth century), Istanbul. (Photograph by the author)



Mosaics, Fig. 3 Hagia Sophia (thirteenth century), Istanbul. (Photograph by the author)

angles so as to reflect light as effectively as possible. Seen from a distance, the singular mosaic pieces appear to model figurative forms. In the present time, this technique is considered as somewhat comparable to the compositions of the Neo-impressionist painters of the 1880s. Thus, the mosaic masters possessed knowledge of the eye's

ability to apprehend color mixing. They also acquired understanding of the changes colors underwent with distance and also of the interaction of light and material.

The images in the Byzantine basilicas are often constructed from contrasting colors. Red stands against green in costumes and clothing designs. A pair of red shoes usually meets green grass. When two contrasting colors meet, there is an increase in both colors. Another approach in the sixth century, used by the mosaic masters to avoid changes in color, was to apply a contour line between two colors. This contour line prevents the colors from spilling over into each other. Instead, the colors are depicted without visible distortion. The gold mosaic primarily represented the halos of the holy figures but was also placed in the background. The subject of the church mosaics was to serve the spirit, not the body. In the images, all suggestions of movement have been avoided; instead, an "eternal existence" is shown. The combined effect in the milieu of the church is a glittering, multicolored immaterial curtain. The material, the light, and the surface interact.

By studying the mosaics, it is possible to see how the mosaic masters applied the tessera pieces with skill and knowledge of color and what happens optically at a distance. The mosaic masters mixed colors optically for entire large areas of color. For instance, in order to give a lawn its various shades of green and thus get a nuanced and living surface, the mosaic pieces were placed against each other in turquoise, green, and gray-green for the purpose of combining and so forming the greenness of the grass. Gold mosaic was also used – in addition to create areas of gold color and point out important details – to mix the colors in larger fields.

When viewing a mosaic wall, the changes of the material from matte to shiny can be discerned. How the mosaic is reflected depends on its angle against the wall. Two pieces of the same inherent color placed with different tilt angles result in that each one will have a different appearance. This effect results from light projecting and reflecting different information to the eye. Reflection is also dependent on how the light strikes the mosaic pieces, on whether the tesserae are made of glass (shiny) or marble (often matte), on the viewing



Mosaics, Fig. 4 The Golden Hall, Stockholm City Hall (1923). (Photograph by the author)

angle, and on how the viewer moves in the room. No movement of the subjects or the motifs on the walls is discernible. The hues are perceived as stable and unchangeable, and unaffected by nearby colors. But with the movement of the viewer in the room, and the changes in the eye's view, the mosaic walls are transformed from matte to shiny, and back again. Through the mosaic, the viewer participates in the process of change.

In the present time, it is possible to find the impact of late Antique mosaics in Sweden, for example. In 1919, the Swedish artist Einar Forseth made a study trip to Italy. The architect Ragnar Östberg designed the Stockholm City Hall in 1923, and Einar Forseth was appointed to create the Golden Hall, one of the main lounges for dining and gathering in the edifice. Forseth chose mosaic for his commission. The Golden Hall is thus covered with more than 18 million glass and gold mosaic pieces (Fig. 4). The mosaics depict Swedish

historical figures and interpret certain events in the Swedish history, but the technique, the imaging, and the colors are in Byzantine style with its origins in Ravenna. Besides the Stockholm City Hall, the Byzantine technique can be seen in an underground station situated in Rinkeby, a suburb of Stockholm. In his adomment from 1975, the artist Nisse Zetterberg used mosaic in gold, in the tesserae technique, applied with different angles to the background. The mosaics depict patterns, symbols, and runes inspired by the Viking age, once found in the Rinkeby excavations.

Cross-References

- [Appearance](#)
- [Color Contrast](#)
- [Neo-impressionism](#)
- [Unique Hues](#)

Motion and Color Cognition

Donald Hoffman

Department of Cognitive Sciences, University of California at Irvine, Irvine, CA, USA

Synonyms

Achromatopsia; Color from motion; Dynamic color spreading; Neon color spreading

Definition

Motion and color cognition is a branch of cognitive neuroscience devoted to using brain imaging, psychophysical experiments, and computational modeling to understand the interactions between motion and color processing in the human visual system.

Overview

The relationships between the perceptions of color and visual motion are complex and intriguing. Evidence from patients with certain rare forms of damage to the cerebral cortex indicates that color and motion are to some extent processed separately within the visual system. This conclusion is buttressed by physiological evidence for separate neural pathways for the processing of color and motion and psychophysical evidence that the perception of motion in color displays is greatly reduced when the colors are isoluminant. However, experiments and demonstrations reveal that the perceptions of color and motion can interact. Motion can trigger the perception of color in achromatic stimuli and in uncolored regions of space. Color can determine the perceived direction of motion, and changes in color can trigger the perception of apparent motion. This article briefly reviews the evidence for separate processing of color and visual motion and for the ways they interact.

Evidence for Separate Processing of Color and Motion

Patients with the rare condition of *cerebral achromatopsia* cannot perceive colors, due to damage of cortex in the inferior temporal lobe of both cerebral hemispheres [1, 2]. Cerebral achromatopsia differs from the more common condition of color blindness, in which one or more of the normal three cone systems in the retinas are missing. A color-blind person missing the long-wavelength cone system, for instance, can still perceive many colors but has a range of color perceptions more limited than that of a normal trichromat, who has all three cone systems. Cerebral achromatopsics can have all three cones systems intact in their retinas but are still unable to perceive any colors because of damage to visual cortex. Yet many cerebral achromatopsics have no difficulty perceiving visual motion, indicating that the processing of color within the human visual system can be damaged without also damaging the processing of visual motion. (See also the Encyclopedia of Color Science and Technology entry on color naming and categorization in inherited color deficiencies and the entry on color perception and environmental impairments.)

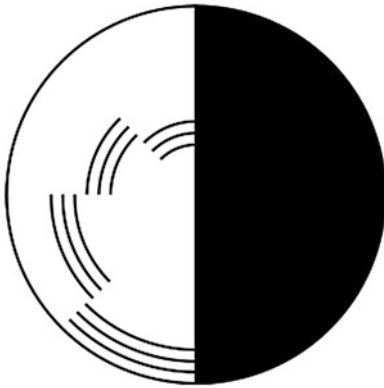
Conversely, patients with the rare condition of *cerebral akinetopsia* cannot perceive motion, due to damage of cortex in the middle temporal area of both cerebral hemispheres [2, 3]. According to Josef Zihl and his colleagues, one such patient “. . . had difficulty, for example, in pouring tea or coffee into a cup because the fluid appeared to be frozen, like a glacier. In addition, she could not stop pouring at the right time since she was unable to perceive the movement in the cup (or a pot) when the fluid rose” [2, 4]. Yet this patient was able to perceive color, indicating that the processing of motion within the human visual system can be damaged without also damaging the processing of color.

Cerebral achromatopsia and cerebral akinetopsia indicate that color and motion are to some extent processed separately within the visual system. Further evidence for their separate processing comes from neuroanatomical and neurophysiological studies revealing that the visual

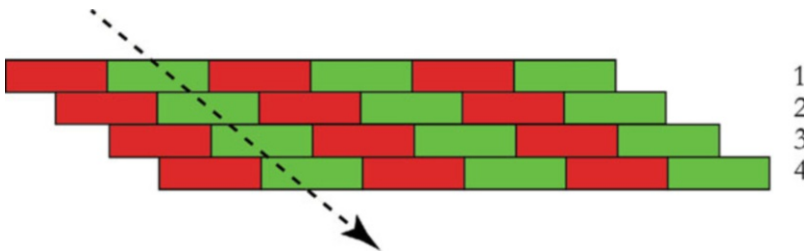
system has a parvocellular pathway that processes color and detailed form information and a magnocellular pathway that processes motion, depth, and coarse form information [5]. These pathways are clearly segregated in the lateral geniculate nucleus of the thalamus, a relay station for visual information traveling from the eyes to visual cortex. One psychophysical correlate of this separate processing is the observation that perceived motion in color displays is greatly reduced when the displays are made isoluminant, an effect first discovered in 1911 by Pleikart Stumpf [6].

Evidence for Interactions in the Processing of Color and Motion

However, despite the anatomical, neuropsychological, and psychophysical evidence for the



Motion and Color Cognition, Fig. 1 Benham's top. Spinning this achromatic pattern leads to perceived colors



Motion and Color Cognition, Fig. 2 Color and motion correspondence. This figure depicts four frames of a movie in which the colored blocks could, in principle, be

perceived either as moving to the left or as moving to the right. Color information, even at isoluminance, biases us to see movement to the right

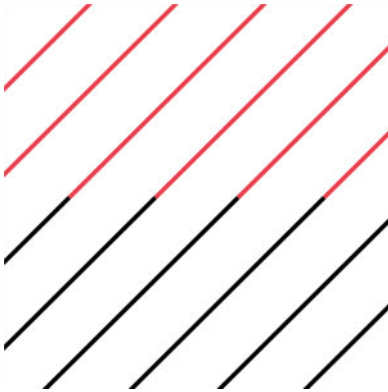
separate processing of motion and color, there are several intriguing visual phenomena which clearly demonstrate that the perception of motion and color can interact. An early example was discovered by the French monk Benedict Prevost in 1826, when he noticed that if he waved his hands in the dim light of the cloisters, then colors appeared near his moving fingers [7].

Charles Benham marketed a top in 1895 with a pattern similar to that shown in Fig. 1 [8]. Spinning the top leads to the perception of the colors of an artificial spectrum, again demonstrating an interaction between color and motion. (See also the Encyclopedia of Color Science and Technology entry on illusory colors.)

Karen Dobkins and Tom Albright discovered in 1993 that color can influence the perceived direction of motion [2]. They showed subjects a movie in which a row of red and green blocks appeared to move horizontally. Four frames from such a movie are shown schematically in Fig. 2. From one frame to the next, the blocks moved exactly one half of a block width, so that the movie is compatible, in principle, with being seen as blocks moving to the left or to the right. However, if the brightness of the green blocks differs from that of the red blocks, then the visual system sees a direction of motion consistent with this, i.e., such that the brightness of each block does not appear to change as it moves. However, if all the blocks are equal in perceived brightness, the visual system sees a direction of motion consistent with the colors, i.e., such that the color of each block does not appear to change as it moves. Thus, color information by itself, apart from luminance information, can influence the perceived

direction of motion. Similar evidence for the interaction of color and motion comes from experiments showing that color can affect the perceived direction of discrete apparent motion stimuli [9] and of sine-wave gratings [10, 11].

Conversely, V.S. Ramachandran discovered in 1987 that motion can “capture” color borders. If an illusory contour or a group of random dots are moved in the vicinity of a color border, then that border appears to move in the same direction as the illusory contour or random dots [12].

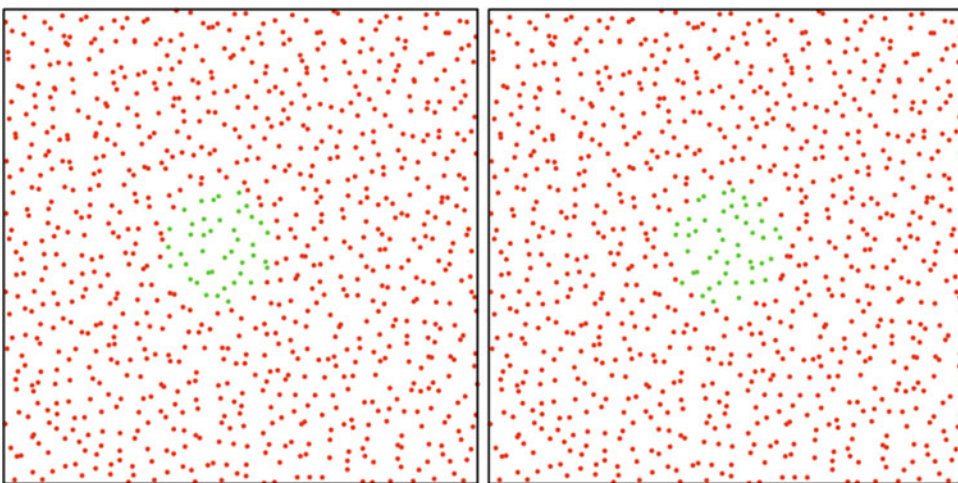


Motion and Color Cognition, Fig. 3 Wallach’s display. The *red neon* color spreading seen in the upper half of the *square* becomes more pronounced when the *lines* are moved *horizontally*, but appear to be moving *vertically*

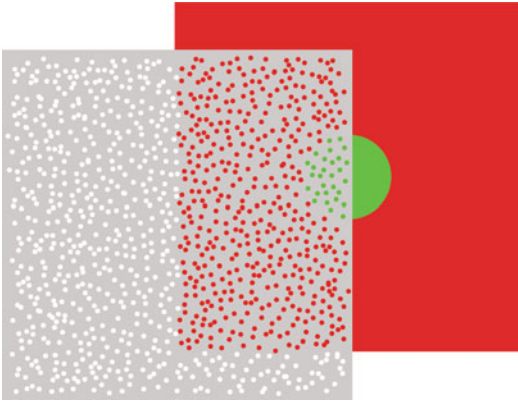
Hans Wallach in 1935 created a display similar to that shown in Fig. 3 [13]. As can be seen in the figure, the oblique lines are red in the upper half of the display and black in the lower half. The red color of the lines in the upper half appears to spread faintly through the white regions, a phenomenon known as neon color spreading.

Wallach then put the oblique lines in motion, translating them sideways underneath a rectangular window, indicated in Fig. 3 by a gray frame. This leads to a bistable perception of the motion of the lines. Sometimes observers see them translating horizontally, as in fact they are. But sometimes, due to the ambiguities of motion seen through small apertures, observers see the lines translating vertically. In this latter case, observers report that the red neon color spreading is greatly enhanced. Thus motion can interact with color by enhancing neon color spreading.

A similar interaction between color and motion can be seen in displays of dynamic color spreading [2, 14]. Two frames from such a display are shown in Fig. 4. Each frame has a few hundred dots, which do not change position from frame to frame. Some of the dots are colored red and others, within a circular region, are colored green. The circular region in which dots are colored green moves slightly from frame to frame. This leads to the perception of a glowing green



Motion and Color Cognition, Fig. 4 Two frames from a movie display of dynamic color spreading. These frames can be stereo fused to get an impression of a glowing *green* disk hovering over a field of *red* dots



Motion and Color Cognition, Fig. 5 Dynamic color spreading seen amodally. The *red dots* appear to be holes in a sheet of paper, and through the holes one sees a *green* disk moving over a *red* background

disk that moves over a field of static red dots. One can get feel for the perception of the glowing disk by using Fig. 4 as a stereo display, and cross fusing the two images. One sees a glowing green disk floating in front of the field of red dots.

Sometimes the green disk does not appear to float in front of a field of red dots. Instead, the red dots appear to be holes in a white sheet of paper, and through the holes one sees a solid green disk moving over a uniformly colored red background. This “amodal” perception of dynamic color spreading is illustrated in Fig. 5. One can get a feel for this amodal perception by again treating Fig. 4 as a stereo display, but this time fusing the two images by diverging the eyes rather than crossing them.

Displays of dynamic color spreading demonstrate that apparent motion can trigger the creation of illusory objects and the spreading of color over the surfaces of such illusory objects. The surface color can have a neon quality if the object is seen floating in front of the dots or a paint quality if the object is seen amodally.

Summary

Evidence from neuropsychology, neurophysiology, and psychophysics indicates that color and motion are to some extent processed in separate

pathways of the visual system, but that these visual pathways interact in a sophisticated fashion, and that together they can create the appearance of objects in motion with colored surfaces.

Cross-References

- [Color Phenomenology](#)
- [Color Processing, Cortical](#)
- [Color Psychology](#)
- [Subjective Colors](#)

References

1. Sacks, O.: *An Anthropologist on Mars*. Vintage Books, New York (1995)
2. Hoffman, D.D.: *Visual Intelligence: how we create what we see*. W.W. Norton, New York (2000)
3. Zeki, S.: Cerebral akinetopsia (visual motion blindness): a review. *Brain*. **114**, 811–824 (1991)
4. Zihl, J., von Cramon, D., Mai, N.: Selective disturbance of movement vision after bilateral brain damage. *Brain*. **106**, 313–340 (1983)
5. Carlson, N.J.: *Physiology of Behavior*, 10th edn. Allyn & Bacon, New York (2009)
6. Todorovic, D.: A gem from the past: Pleikart Stumpf’s (1911) anticipation of the aperture problem, Reichardt detectors, and perceived motion loss at equiluminance. *Perception*. **25**, 1235–1242 (1996)
7. Cohen, J., Gordon, D.A.: The Prevost-Fechner-Benham subjective colors. *Psychol. Bull.* **46**, 97–136 (1949)
8. von Campenhausen, C., Schramme, J.: 100 years of Benham’s top in colour science. *Perception*. **24**(6), 695–717 (1995)
9. Papathomas, T.V., Gorea, A., Julesz, B.: Two carriers for motion perception: color and luminance. *Vis. Res.* **31**, 1883–1891 (1991)
10. Mullen, K.T., Boulton, J.C.: Absence of smooth motion perception in color vision. *Vis. Res.* **32**, 483–488 (1992)
11. Chichilnisky, E.J., Heeger, D., Wandell, B.A.: Functional segregation of color and motion perception examined in motion nulling. *Vis. Res.* **33**, 2113–2125 (1993)
12. Ramachandran, V.S.: Interaction between colour and motion in human vision. *Nature*. **328**, 645–647 (1987)
13. Wallach, H.: Ueber Visuell Wahrgenommene Bewegungsrichtung. *Psychol. Forsch.* **20**, 325–380 (1935)
14. Hoffman, D.D.: The interaction of colour and motion. In: Heyer, D., Mausfeld, R. (eds.) *Colour Perception: Mind and the Physical World*, pp. 361–377. Oxford University Press, Oxford (2003)

mRGCs

- [Melanopsin Retinal Ganglion Cells](#)

Multi-beam

- [Automotive Front Lighting System \(Status 2020\)](#)

Multispectral Imaging

- [Hyperspectral Imaging](#)

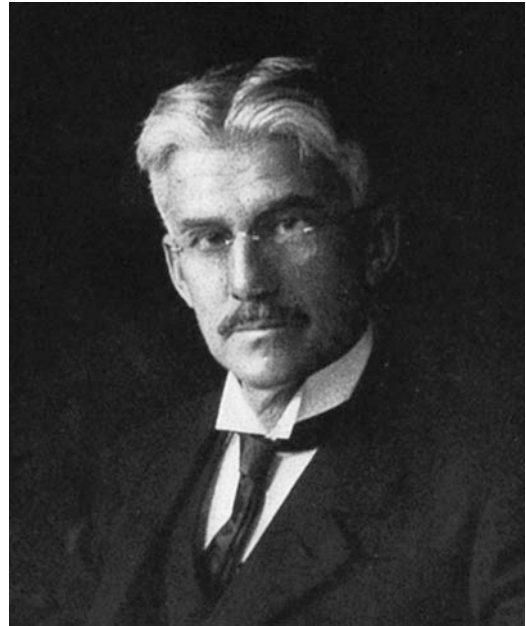
Munsell, Albert Henry

Rolf G. Kuehni
Charlotte, NC, USA

Biography

Munsell was born on January 6, 1858, in Boston, MA, where his father was in the piano business. After high school he attended the Massachusetts Normal Art School in Boston. In 1879 he studied Ogden Nicholas Rood's influential book *Modern Chromatics*. In 1881 he was named an instructor and later a lecturer at this school, positions he held for 25 years. He was awarded a scholarship that made it possible for him to study from 1885 to 1888 at the Académie Julian and the École nationale supérieure des Beaux-Arts in Paris and 1 year in Rome. After he returned he was an active painter of portraits and seascapes. In 1889 he received a patent for an adjustable artist's easel. In 1894 he married Julia Orr, the daughter of a New York financier, with whom he had a son,

Albert Ector Orr Munsell, and three daughters. Munsell died on July 28, 1918, in Brookline, MA [1].



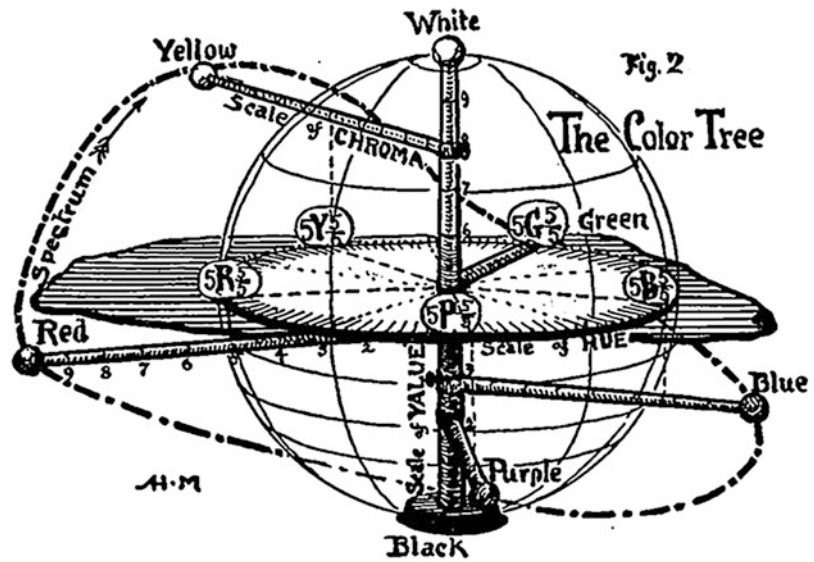
Photograph from *A Color Notation* [5]

M

Major Accomplishments/Contributions

As a lecturer at the art school, Munsell became concerned with how to teach students about colors in a meaningful manner. In 1899 he developed a model of a balanced color sphere that, when spun rapidly, resulted in a gray appearance, thus showing a kind of color balance. He received a patent for it in 1900 [2]. Munsell met O. N. Rood, then professor of physics in New York, in 1899 showing him his color sphere, with Rood commenting positively on it. Simultaneously, Munsell concerned himself with the design of the interior of the color sphere. In April 1900 he sketched a hue circle based on the decimal system, with five primary, five secondary, and ten intermediate hues. It continues to be the basis of the modern system. Munsell, from his studies in painting, had a clear idea of lightness or value, as it was typically called in painters' circles. In 1901 he

Munsell, Albert Henry,
Fig. 1 Munsell's
 schematic depiction of the
 color tree



obtained a patent for a photometer he called a luminometer. It was manufactured in small numbers in the following years [3]. The basic question of using a logarithmic or a power root scale to relate the physical data of the luminometer to perceived lightness in his system was a question not fully resolved until after Munsell's death. In 1902 Munsell hand-plotted color intensity and lightness data for different pigments, earlier established by W. Abney, and realized that a sphere was an incomplete representation of perceived colors. As a result he began to name his color solid a "color tree." Having established a hue circle and a subjective lightness scale, he was aware that for completeness he needed to define a chromatic intensity scale. It is not established how he arrived at the term "chroma" for that purpose. In March 1902 he sketched a color solid based on the three attributes as shown in more detail in Fig. 1 (Ref. 4, 1907). Samples were to be spaced according to perceptual equal differences by parameter. Munsell began to establish such samples to fill limited hue pages. Munsell described the system in *A Color Notation*, first published in 1905 [5]. A first edition of the *Atlas of the Munsell Color System* was published in 1907 containing eight charts of painted samples of ten different hues [6]. In the 1915 edition of the atlas, the number of samples was doubled. Munsell also introduced artist's tools based on his system: Munsell crayons and Munsell

water colors. As a painter Munsell was interested in color harmony, and he established nine principles based on his color solid [7].

Throughout the development of the system, he consulted with a broad group of scientists and artists and gave many presentations in the USA and in Europe. In 1917 he formed the Munsell Color Company to operate the business of producing the atlas. After his passing it was taken over by his son A. E. O. Munsell and other stockholders. At about the same time, the National Bureau of Standards began to show interest in the system and supported sample measurement and expansion of the system. An enlarged edition with 20 hues was published in 1929 [8]. In the 1940s extensive experiments were made under the auspices of the Optical Society of America resulting in the Munsell Renotations (for Munsell system colorimetric renotation data, see, e.g., Ref. [9]), colorimetric definitions of a revised version of samples of the atlas that continue to be the basis of the modern system. Munsell's is perhaps the most important color atlas system yet developed.

References

1. Munsell, A. H.: The National Cyclopaedia of American Biography 36, p. 316 (1950)
2. Munsell, A. H.: Color sphere and mount. US Patent No. 640,792 of 9 Jan 1900

3. Munsell, A.: Notebooks 1900–1918. Accessible at www.cis.rit.edu/research/mcsl2/online/munselljournals.php. Accessed 15 Jan 2014
4. Munsell, A.H.: Color and an Eye to Discern It. Self-Published, Boston (1907)
5. Munsell, A.H.: A Color Notation. Ellis, Boston (1905)
6. Munsell, A.H.: Atlas of the Munsell Color System. Ellis, Boston (1907)
7. Munsell, A.H., Cleland, T.M.: A Grammar of Color. Strathmore Paper, Molineage (1921)
8. Berns, R.S., Billmeyer, F.W.: Development of the 1929 Munsell book of color. Color Res. Appl. **10**, 245–250 (1985)
9. Wyszecki, G., Stiles, W.S.: Color Science, 2d edn. Wiley, Hoboken (1982)

Further Reading

- Kuehni, R.G.: The early development of the Munsell system. Color Res. Appl. **27**, 20–27 (2002)
- Nickerson, D.: History of the Munsell color system and its scientific application. J. Opt. Soc. Am. **30**, 575–585 (1940)

N

Natural Color Distribution

► [Color Scene Statistics, Chromatic Scene Statistics](#)

Natural Colorants

► [Colorant, Natural](#)

Natural Dyes

► [Colorant, Natural](#)

Natural Hair Colorant

► [Melanin](#)

Natural Hair Pigmentation

► [Melanin](#)

Natural Object

► [Blackbody and Blackbody Radiation](#)

Natural Scene Statistics

► [Color Scene Statistics, Chromatic Scene Statistics](#)

Neo-impressionism

Georges Roque
Paris, France

Definition

Neo-impressionism is an artistic movement coming after Impressionism. It started as a section of the eighth and last Impressionist Exhibition in 1886. Its founding figure was Georges Seurat (1859–1891), whose monumental painting, *A Sunday on the Grande Jatte*, reworked for its presentation in the 1886 exhibition, is a kind of standard bearer of the whole movement. Seurat died in 1891 at his early thirties. His death corresponds more or less to the end of Neo-impressionism first life, all the more since Camille Pissarro gave up the new technique around the same period. After Seurat's death, Paul Signac (1863–1935) became the leader of the movement and his chief publicist. At the beginning of the

Georges Roque has retired.

twentieth century, Neo-impressionism had a revival during the Fauve period; most of the pioneers of abstract art also had a Neo-impressionist phase in their way to abstraction. The main features of Neo-impressionism are a faith in science and color science and the use of bright colors and of a special technique (optical mixture) aimed at giving more luminosity to colors; this technique, which implies a mechanical application of the brushstroke, was also intended to suppress the skill of the hand as characteristics of the individualism dear to the Impressionists.

Overview

The term “Neo-impressionism” was coined by the critic Félix Fénéon, who was the earliest and one of the best spokesmen of the movement. The Neo-impressionist artists did not want to break with the Impressionists (one of them, Camille Pissarro, was himself one of the leading Impressionist painters) but continue it in a more rigorous and scientific way. However, one of the differences between the two movements is that, unlike the Impressionists, the Neo-impressionists wanted to render permanent features and not ever-changing moments. A superficial characteristic of their brushstroke is the use of small dots of color, hence the term of pointillism in order to describe their technique.

The main firsthand sources of information to understand Neo-impressionism are (a) an unpublished letter of Seurat to Maurice Beaubourg (1890) of which different drafts are known ([1], pp. 113–114), (b) a letter of Pissarro to his dealer Durand-Ruel (1886) ([1], pp. 54–55), (c) the book written by Signac in order to champion the movement (1899) [2], and (d) an article written by the critic Félix Fénéon and based on information provided by the artists (1886) ([1], pp. 108–110); this article is sometimes considered as the Neo-impressionist manifesto.

Art and Science

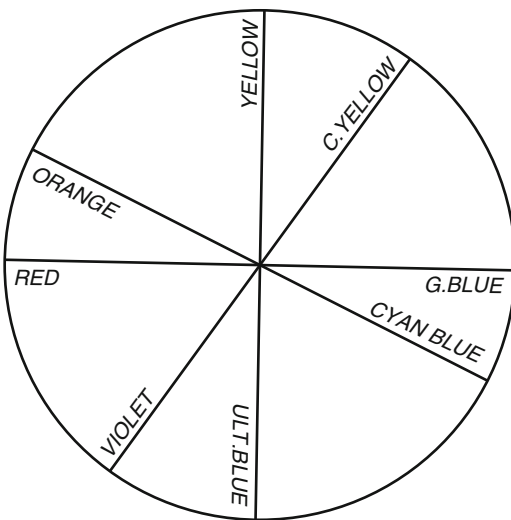
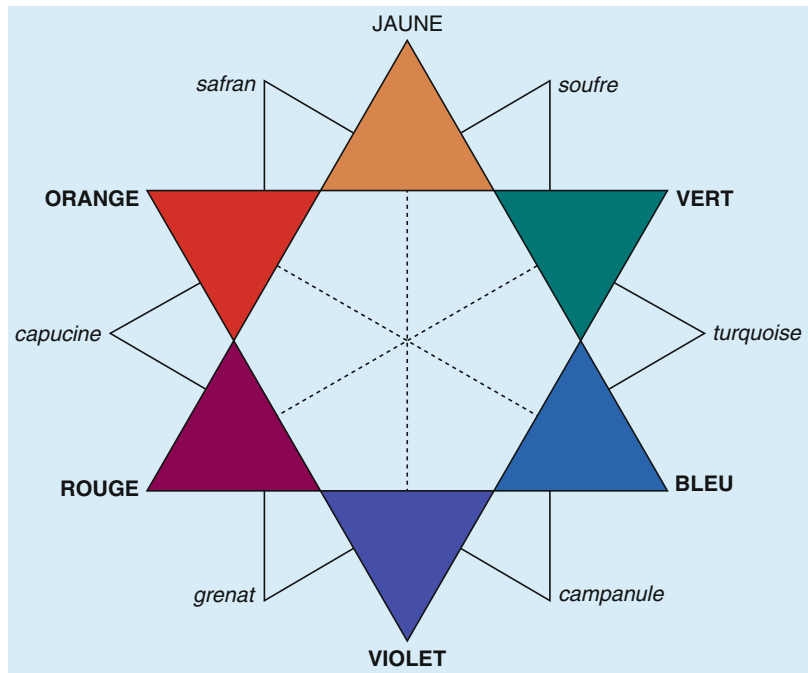
According to these writings, the main sources for Neo-impressionist color practice are Michel

Eugène Chevreul [3], basically known through Charles Blanc [4], and Ogden Rood [5]. Now these sources are contradictory, at least concerning color harmony, as Chevreul used pairs of complementary colors popularized through Blanc’s color star (Fig. 1), while Rood used new pairs of complementary colors based on Helmholtz’s researches (Fig. 2). This explains Seurat’s hesitation between the two color systems.

The Neo-impressionists used scientific sources and were very interested in what science could bring them; for this reason, they visited Chevreul on one occasion and his assistant on another, in order to understand better the “division of light” (on the importance of Chevreul for them, see [6], pp. 350–377). Their interest for color science raises the issue of whether or not they had a scientific approach to color. This is a very debated topic, as answers have been completely opposed: some specialists think Seurat and his fellow painters used a “scientific” method based on color science; on the contrary, others consider that the Neo-impressionist artists misunderstood the scientists’ advices, so that their practice was pseudo-science. Why are there so different positions? It depends on the (implicit or explicit) conception of the relationship between art and science that frames the specialists’ thinking. For those who want to bridge the gap between art and science, Seurat’s practice is obviously applied science. Conversely, those who hold that science remains out of the scope of artists will emphasize the gap between scientific data and artistic practice and will insist accordingly on the artists’ misunderstanding of science. In order to avoid these difficulties, the distinction between color science and color theory might prove useful ([6], pp. 12–17; [7], pp. 44–46). Their function, indeed, is different: color science is a set of experiments and principles aiming at understanding chromatic phenomena and expressing their properties, while color theory is a set of principles, doctrines, rules, and formulas directed at artists and suggesting recipes for choosing, preparing, organizing, and matching colors on the canvas. From this point of view, the difference between an artistic and a scientific approach to color can be more easily understood. For example, Chevreul had

Neo-impressionism,

Fig. 1 *Color star* from Charles Blanc, *Grammaire des arts du dessin*, 1867



Neo-impressionism, Fig. 2 Rood's color wheel, 1879

noticed that two areas of complementary colors tend to enhance each other when juxtaposed ([3], § 237), but he never recommended this effect to be deployed in painting, despite what is usually assumed. In fact, for him, the enhancing effect was not especially desirable. However, for painters preoccupied with luminosity and saturation,

enhancing colors by the juxtaposition of complementary hues soon became an excellent recipe. Examples like this show that “misunderstanding” science is not a mistake that can be avoided by the better-informed artist, but is instead the inescapable result of a difference of aims between the painter and the color scientist.

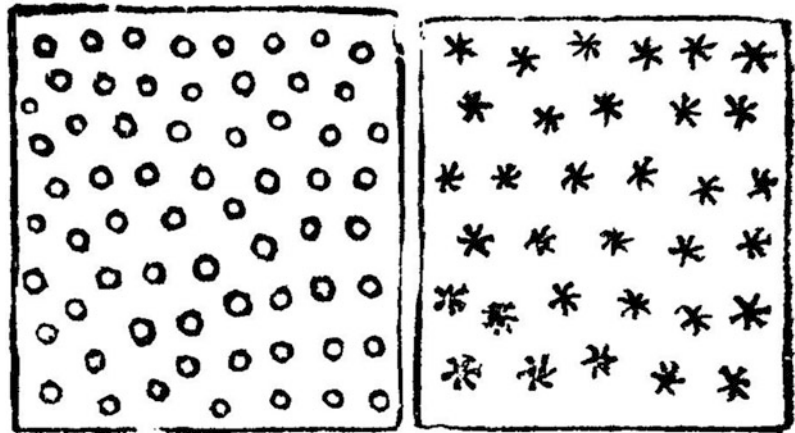
Optical Mixture

Different origins of the so-called pointillism can be traced from the sources given by the artists. The first is Charles Blanc, who coined the expression “optical mixture,” a technique already used by Delacroix, according to him, as well as by the Impressionists. The basic idea is to mix colors not on the palette, but juxtapose them on the canvas, so that when seen at a distance, they fuse into the eye and produce a new color not present at such on the canvas. As Blanc put it:

If at a distance of some steps, we look at a cashmere shawl, we generally perceive tones that are not in the fabric, but which compose themselves at the back of our eye by the effect of reciprocal reactions of one tone upon another. Two colors in juxtaposition or superposed in such or such proportions, that

Neo-impressionism,

Fig. 3 Drawing from Charles Blanc's *Grammaire des arts du dessin*, 1867



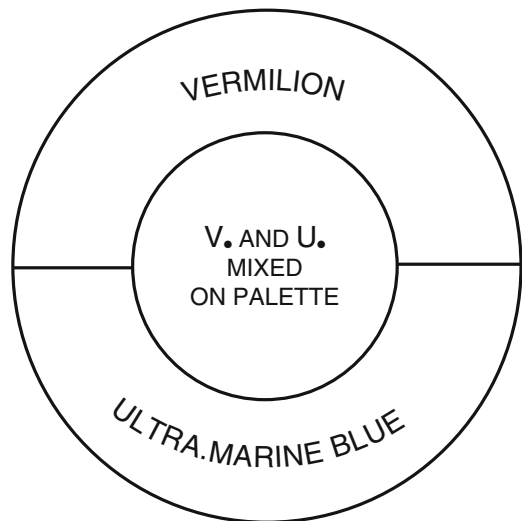
is to say according to the extent each shall occupy, will form a third color that our eye will perceive at a distance, without having been written by weaver or painter. This third color is a resultant that the artist foresaw and which is born of optical mixture. ([4], p. 475)

In order to make his suggestions more concrete, he added a drawing of little dots (Fig. 3, left) or little stars (Fig. 3, right) on a uniform ground: "A similar phenomenon will be produced upon a yellow stuff starred with violet and upon a blue stuff sown with orange spots" ([4], p. 475).

A physicist, Ogden Rood, gave a more complete description of optical mixture in his book published in 1879 and translated into French in 1881. He first described a method given by Mile in 1838 in order to mix colors in the eye from small contiguous dots of colors ([5], p. 140). Rood also quoted John Ruskin, who recommended a similar method ([5], p. 140). It must be noted that Monet as well as Signac praised a lot Ruskin's book *The Elements of Drawing*.

Rood himself gave painters the same advice:

There is, however, another lower degree of gradation which has a peculiar charm of its own, and is very precious in art and nature. The effect referred to take place when different colours are placed side by side in lines or dots, and when viewed at such a distance that the blending is more or less accomplished by the eye of the beholder. Under these circumstances the tints mix on the retina, and produce new colours, which are identical with those which are obtained by the method of revolving disks. ([5], pp. 279–280)



Neo-impressionism, Fig. 4 Drawing of experiment from Rood's *Modern Chromatics*, 1879

Luminosity

The end of the quote explains why the Neo-impressionists were so interested by the method suggested by Rood (after Blanc and Ruskin): Rood made a fascinating experiment in order to compare the luminosity of colors mixed on the palette with the optical mixture of the same hues. For this purpose, he used two rotating disks. The larger disk was divided into two equal halves, painted with vermilion in the upper part and ultra-marine blue in the lower part (Fig. 4).

Neo-impressionism,

Fig. 5 Compared result of color mixtures, Rood's *Modern Chromatics*, 1879

Mixture on Palette.	Mixture by Rotation.
50 violet + 50 Hooker's green..	= { 21 violet + 22.5 Hooker's green + 4 vermilion + 52.5 black.
50 violet + 50 gamboge.....	= 54 violet + 20 gamboge + 26 black.
50 violet + 50 green.....	= 50 violet + 18 green + 32 black.
50 violet + 50 Prussian-blue...	= 47 violet + 49 Prussian-blue + 4 black.

Now, an equal number of drops of the same washes of watercolor pigments were mixed on the palette in order to paint the small disk with this mixture. After rotating the two disks, Rood realized that the smaller disk looked gray and dull, while the larger one looked brighter. So, in order to obtain a similar result on both disks, he had to add black to the larger one. But the surprising result of the experiment is that Rood had to add up to 52% of black to the larger disk in order to match the mixture made on the palette (Fig. 5).

Many artists were struck by this experiment, and Fénéon gave it as an example in order to explain the new technique and its superiority in terms of luminosity. Quoting the results of Rood's experiment, Fénéon stressed that "the luminosity of optical mixtures is always superior to that of material mixture" ([1], p. 109) and justified accordingly the new Neo-impressionist technique.

However, the theory of "optical mixture," as understood by painters, when suggesting that it should be possible to get an additive-like mixture by juxtaposing pigments so that they mix in the eye instead of physically, has been a source of enduring confusion. The point that often eluded painters is that this type of optical mixture does not add together the luminous energies of the individual colors; it merely averages them. Furthermore, there is a contradiction in the theory, when painters wished to combine the two principles: complementary colors and optical mixture. In principle, both have in common to heighten colors: they heighten hues in the case of juxtaposing complementary colors; they heighten luminosity in the case of optical mixture. However, painters were not aware of the fact that both principles correspond to two different laws of perception. The color contrast only functions if the samples of color are big enough. For when the juxtaposed zones are thin, what happens is exactly the opposite: instead of enhancing each other, that

is, to exaggerate their difference, they tend to "assimilate," that is, to produce visually a dirty gray.

This leads to a puzzling conclusion: if juxtaposing dots of complementary colors, as the Neo-impressionist painters did, does not give the paintings more luminosity, how could the great luminosity of their paintings be explained? The answer has been given by Robert Herbert, the scholar most familiar with Neo-impressionism. As he explained, if the Neo-impressionist paintings are indeed very luminous, it is because the optical mixture does not work, that is, that the dots are big enough to be still perceived at the normal distance at which we look at the paintings! And it is precisely because the dots do not mix in order to produce new hues that they keep their luminosity ([8], p. 19). It is not the case to stress here that the painters were "wrong" in their use of theory, but only to understand that they were so fascinated by the possibility of obtaining a luminosity similar to that of color-light that they forgot scientists' warnings against the danger of confusing pigments and light colors.

This means that the numerous misunderstandings arising from the notion of optical mixture are inevitable since they are due to the very ambiguity of the concept, which suggests a strong analogy between light and the eye, as opposed to pigment mixture. However, painters work with pigments and not with light but were eager to obtain through their pigments a luminosity similar to that of additive mixture. Hence, the opposition in Signac's writings between mixtures on the palette he considered disgusting and optical mixture "purity."

The Neo-impressionist technique is then rather complex and cannot be limited to "pointillism." The term "pointillism" is often considered as a synonym of Neo-impressionism but is not. Signac insisted a lot on the fact that the dots are just a means and cannot be considered as a characteristic of Neo-impressionism: far more important for

its technique is the “divided touch.” As the painter wrote: “*Division* is a complex system of harmony, an aesthetics rather than a technique. The *dot* is only a means. To divide is to seek the power and harmony of color, through representing colored light by its pure elements and by using the optical mixture of these pure elements, separated and proportioned according to the essential laws of contrast and gradation” ([1], p. 121).

Harmony

As results from this quote, the idea of harmony in Neo-impressionism goes beyond color, due to the Neo-impressionists’ adherence to symbolism. In his letter to Beaubourg in which he explained his main aesthetic ideas, Seurat also related harmony with tones and lines ([1], pp. 113–14). The idea of color harmony must also be associated with the concept of harmony in anarchism, whose ideals were shared by several of the painters. For the anarchists, indeed, the idea of harmony makes it possible to reconcile the individualism characteristic of the movement with collective life. Similarly, a parallel can be drawn between the individuality of colors and the painting as a whole, as all the single elements used (dots) interact with each other and contribute to the general harmony of the painting [9].

Neo-impressionism and Color Development in Twentieth-Century Art

It is striking to note that most of the painters that counted for the development of color in the first half of the twentieth century passed through a Neo-impressionist period. The list is quite impressive: Matisse, Mondrian, Kandinsky, Malevich, Braque, Delaunay, Kupka, and so on. This raises the question of why Neo-impressionism was so important for the development of the practice of color in painting.

The first feature to be noted is the extraordinary homogeneity of the pictorial matter: the dots that serve to render nature are the same as used for a human body. And even in the human body, no

difference is made between hair, skin, clothes, and so on. Everything is treated in the same way; besides color, this is also true for lines and spaces. This homogeneity is important for several reasons. First, it means the rejection of the skill of the hand as an expressive device. For the Neo-impressionists, what matters are the scientific bases of the technique. Furthermore, through the homogeneous construction of the surface, the touch becomes independent of the objects represented on a painting, since the same kind of dot is used on the whole surface of the canvas. From this point of view, the Neo-impressionists have emphasized special means, which are artificial devices that do not resemble nature.

Yet the fact of using artificial means has another consequence: the touch (as a means) becomes independent of the representation of nature, which was the traditional aim of painting. Finally, not only is the divided touch a technique in order to paint through color only, but it also contains a syntax, that is, a way of organizing the contiguous color dots, as well as a complex harmony system. The vocabulary (the dots), the syntax (the divided touch), and the harmony constitute a totality in itself, independently of the world of the objects the painting is supposed to represent. This is the main reason why the next generations of painters were so interested in Neo-impressionism, read avidly Signac’s book, and had a Neo-impressionist phase: the dots were a way of exploring the possibility of painting through colors only, which eventually led to abstract art [10].

Summing up, it must be emphasized that a lot has been written on Neo-impressionism. On color in particular, the interested reader can also refer to G. Roque [7], J. Gage [11], and P. Smith [12], who have developed different aspects related to Seurat, and G. Roque [13], who refers to Signac and color theory.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Chevreul, Michel-Eugène](#)
- [Color Contrast](#)
- [Color Harmony](#)

- [Complementary Colors](#)
- [Impressionism](#)
- [Palette](#)
- [Retinal Art](#)
- [Unique Hues](#)

References

1. Nochlin, L.: Impressionism and Post-impressionism 1874–1904. Sources and Documents. Prentice Hall, Englewood Cliffs (1966)
2. Signac, P.: D'Eugène Delacroix au néo-impressionnisme. Hermann, Paris (1st ed. 1899) (1964); excerpts in English are published in [1], pp. 116–123 (A full English translation has been published as an appendix (pp. 193–286) of Ratliff, F.: Paul Signac and Color in Neo-impressionism. The Rockefeller University Press, New York (1992))
3. Chevreul, M.-E.: De la loi du contraste simultané des couleurs... Pitois-Levrault, Paris (1839) (The latest English translation is: The Principles of Harmony and Contrast of Colors and Their Applications to the Arts (1855). Kessinger Publishing LLC, Whitefish (2009))
4. Blanc, Ch.: Grammaire des arts du dessin. Renouard, Paris (1867) (The chapter on color has been translated into English in Taylor, J.C. (ed.): Nineteenth-Century Theories of Art, pp. 468–479. University of California Press, Berkeley (1989); the quotations in this article refer to this translation)
5. Rood, O.N.: Student's Textbook of Color or Modern Chromatics, with Applications to Art and Industry. D. Appleton and Company, New York (1st ed. 1879) (1899)
6. Roque, G.: Art et science de la couleur: chevreul et les peintres, de Delacroix à l'abstraction, 2nd edn. Gallimard, Paris (2009)
7. Roque, G.: Seurat and color theory. In: Smith, P. (ed.) Seurat Re-Viewed, pp. 43–64. Pennsylvania State University Press, University Park (2009)
8. Herbert, R.L.: Neo-impressionism. Exhib. cat. Guggenheim Museum, New York (1968)
9. Roque, G.: Harmonie des couleurs, harmonie sociale. 48/14. La revue du musée d'Orsay 12, 62–73 (2001)
10. Franz, E. (ed.): Signac et la libération de la couleur de Matisse à Mondrian. Exhib. cat. Musée de Grenoble, Grenoble (1977)
11. Gage, J.: The technique of Seurat – a reappraisal. In: Gage, J. (ed.) Color and Meaning. Art, Science and Symbolism, pp. 209–218. Thames and Hudson, London (1999)
12. Smith, P.: Seurat and the Avant-Garde. Yale University Press, New Haven/London (1997)
13. Roque, G.: Signac et la théorie de la couleur. In: Ferretti Bocquillon, M. (ed.) Signac. Les couleurs de l'eau. Exhib. cat, pp. 30–39. Musée des Impressionismes and Gallimard, Giverny/Paris (2013)

Neon Color Spreading

- [Motion and Color Cognition](#)

Neon Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Glow discharge lamps](#); [Low-pressure glow discharge lamps](#)

Definition

Lamps that produce light as a result of an electrical glow discharge, generated between two electrodes, in a low-pressure neon, argon, krypton, helium, xenon, or carbon dioxide vapor, which is contained in a transparent bulb or tube.

Note that also glow discharge lamps using no neon but one of the other vapors mentioned in the definition are referred to as neon lamps.

Neon lamps are usually not used for general lighting purposes, and therefore only the fundamentals of their working principle and of their properties are described in this encyclopedia.

Working Principle of Glow Discharge Lamps

In a glow discharge, the electrons taking part in the initial discharge are made free by external sources like thermal collisions between atoms, ultraviolet, or cosmic radiation. In a discharge bulb or tube with gas at very low pressure and a voltage of several hundred volts applied across cold electrodes, the electrodes made free by the external sources move at high speed toward the anode and ionize some gas atoms creating positively charged ions and more free electrons. Some

of the positive ions hit the cathode at high speed, thanks to the low pressure of the gas of less than 0.05 atmosphere. They release by their impact-free electrons from the cathode surface. Electron production is thus not by external means like heated cathodes but by secondary emission as a result of positive ions hitting the cathode surface. Atoms can also be excited to a higher energy state by the moving electrons and ions, and in releasing the extra energy so obtained, photons are created that are responsible for the typical soft and transparent luminous glow. The color characteristics of the glow light are dependent on the type of gas used.

The glow discharge process is only stable at small current densities in the order of milliamperes. By increasing the voltage, the current is increased, and the process will change into that of an arc discharge that is typical for most discharge lamps used for lighting: low- and high-pressure sodium, mercury, and metal-halide lamps. Some arc discharge lamps make use of the glow light process to facilitate ignition of the lamp (the red neon glow light of a starting low-pressure sodium lamp). Glow light starters (used in some fluorescent lamp systems) also make use of this process.

Tubular Glow Discharge Lamps

Tubular neon lamps can be shaped into any form and produced in many different color versions. They are therefore particularly well suited for advertising signs. LED strings now offer an alternative. Tubular neon signs are also popular in decorative art.

Construction

Lead glass is usually used for the tube because it can easily be bent and fused. Sometimes red- or yellow-tinted glass is used to adapt the original color of the glow light. Yellow-tinted glass will, for example, change the color of blue glow light into green.

The length of the tube usually is less than 2 m, and its diameter varies between some 7 and 35 mm. Larger signs are constructed from various pieces of individual tubes.

The electrodes are large rings of pure iron with a diameter of 15–20 mm and a length of about 80 mm.

Pure neon as fill gas results in red light; most other colors can be produced with argon-mercury mixtures. Helium results in warm white light.

Fluorescent powders on the inner wall are often used to extend the range of colors or to improve light output.

Properties

System Luminous Efficacy

The luminous efficacy of clear glass tubes is low with some 5 lm/W. Fluorescent-covered tubes have efficacies depending on the color between some 10 lm/W for red, 30 lm/W for blue, more than 80 lm/W for green, and 75 lm/W for white. The values vary quite a bit with the type of fluorescent powder used.

Lumen Package Range

The luminous flux of neon tubes is expressed in lumen per meter of tube length. Values vary according to the type of gas, type of fluorescent powder (and thus color), lamp current, and tube diameter between 100 lm/m (clear glass neon-filled) and 1600 lm/m (fluorescent-coated tube producing green light).

Color Characteristics

Applying different fill gasses, different tinted glass tubes, and different fluorescent powders enable the production of all the colors of the spectrum.

Lamp Life

Neon tubes have electrically a very long life. Their practical life is often indicated as the time after which the light output has dropped to 50%: 10,000–20,000 h.

Run-Up and Reignition

Ignition of the cold lamp is instantaneous, and hot reignition is immediate as well. This makes it possible to produce flashing neon sign installations.

Compact Glow Discharge Lamps

Neon compact glow lamps were used as indicator lamps on all kinds of electrical and electronic equipment. LED signal lamps have taken over this application. They were also, and still are, used for low-level lighting in night lamps in children's and hospital bedrooms and corridors. Also, here LEDs are an alternative. The lamp's electrodes can be shaped in many different forms so that the glow light can be used to display letters, numbers, and symbols. By making the lamp flashing, a flickering neon glow, for example, in the form of a candle, flame can be produced.

Construction

Compact glow discharge lamps consist of a compact glass bulb with two electrodes closely positioned together. The nickel or iron electrodes can take any form like strip, disk, ring, cross, etc. The filling gas usually is neon giving an orange-red glow. Fluorescent coating is used for green-emitting lamps. Miniature indicator lamps in sizes from some 5 to 15 mm operate on a current of 0.5–1 mA. The operating current of the larger types in the size of an incandescent bulb goes up to 50 mA.

Properties

System Luminous Efficacy

Luminous efficacy is 0.3–1 lm/W, which for indicator lamps is reasonable enough.

Lumen Package

The lumen output, less to much less than one lumen, is low but for indicator lamp applications sufficient.

Color Characteristics

Clear glass bulbs produce orange-red light. With fluorescent coating of the bulb, many more colors can be produced, especially also green, which for indicator lamps is indispensable.

Lamp Life

Lamp life is 10,000–100,000 h.

Run-Up and Reignition

Cold start and reignition are both instantaneous, which of course for indicator lamps is the first requirement.

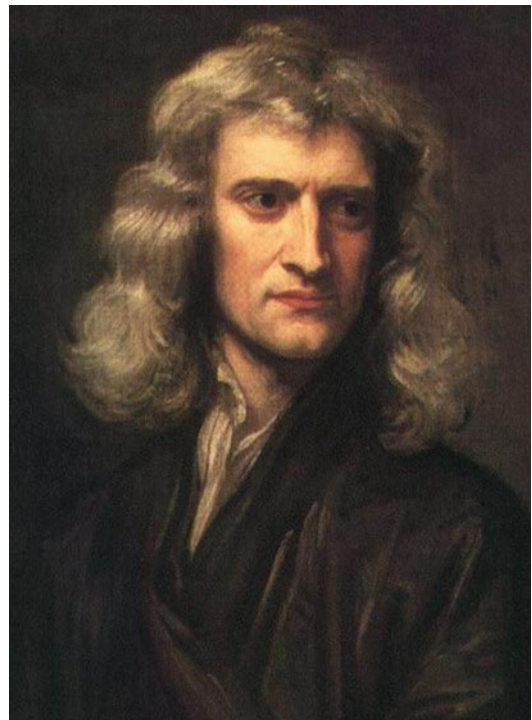
Cross-References

- ▶ [Light-Emitting Diode, LED](#)
- ▶ [Phosphors and Fluorescent Powders](#)

Newton, (Sir) Isaac

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA



1642/1643–1726/1727 (The difference is due to the use and report of the date in two calendars: Julian and Gregorian. According to the Gregorian calendar, the date of Newton's birth was 4 Jan, 1643 and that of his death 31 March 1727) (Portrait of Isaac Newton in 1689 (age 46) by Godfrey Kneller, Wikimedia)

Biography

Isaac Newton was an English physicist and mathematician, who made seminal contributions to several domains of science, and was considered a leading scientist of his era and one of the most influential scientists of all time. He was born prematurely on 25 December 1642 in Lincolnshire, England, of Hannah Ayscough and Isaac Newton (father) [1]. His father passed away before he was born. His mother remarried when he was 3 years old and he was left in the care of his grandmother, a situation he resented.

He attended the King's School in Grantham where he learned Latin among other things until the age of 17. In 1661, he was admitted to Trinity College, Cambridge, and was educated in Aristotelian philosophy. However, Newton also read the works of Descartes, Galileo, and Kepler. In 1665/1666 he spent most of the time at his ancestral place in Lincolnshire because of the dangerous spread of plague in Cambridge. During that time he did most of his experimental work with glass prisms and much of his mental work that resulted in the publication of *Principia* in 1687.

In 1667 he returned to Cambridge and became a fellow of the College of the Holy and Undivided Trinity [2, 3]. In 1669, and at the age of 26, Newton became the Lucasian professor of mathematics. According to his secretary Humphrey Newton (no relation), his lectures were often poorly attended and few understood him and that sometimes he read to the walls [4]. Newton occasionally traveled to London to attend the Royal Society lectures and was named a fellow in 1672. He became its president from 1703 to 1727. Newton was given various levels of support by the Royal Society. His interpretation of the optical experiments was strongly disputed by Robert Hooke, an employee of the society since 1664, as a result of which he published his book *Opticks* only in 1704, after Hooke's death. He was supported by the society in his bitter and controversial dispute with the German polymath and philosopher Gottfried Wilhelm Leibniz over who had developed calculus first [Leibniz's notations are used today].

Newton also dwelt in politics and was a member of the House of Commons between 1689 and 1690 and then again from 1701 to 1702. In 1705 he was knighted by Queen Anne during her visit to Trinity College. He held two government offices: first he was the Warden of the Mint from 1696 to 1700 and then Master of the Mint from 1700 until his death in 1727. At the time of Newton's funeral, the French philosopher Voltaire who was in England compared Newton to Descartes and said of Newton that "he was never sensible to any passion, [and] was not subject to the common frailties of mankind, nor had any commerce with women" [5]. Newton had strong opinions on religion and wrote a number of works, not published during his lifetime, that would have then been considered heretic in that period. Newton never married and died intestate in Kensington, London, when his relatives quarreled over the division of his considerable estate. He is buried in Westminster Abbey in London, England.

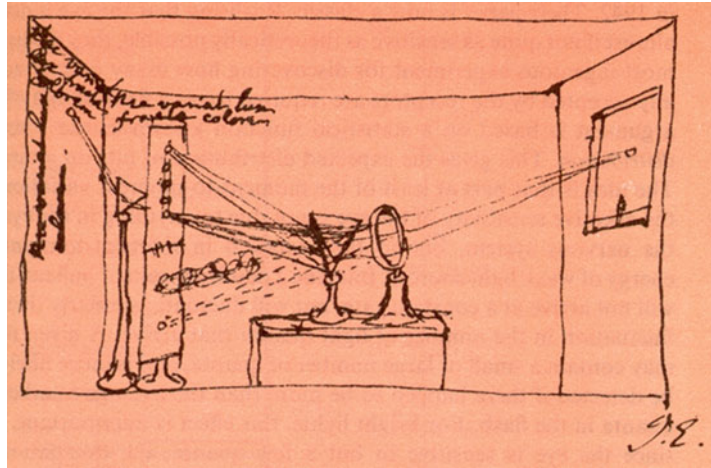
Major Accomplishments/Contributions

Newton's Theory of Color

Arguably our modern understanding of light and color begins with Newton's discovery of light dispersion which he published in 1672. In the late 1660s, Newton started experimenting with the phenomenon of colors and lectured on optics [6]. At the time, it was generally thought that colors were mixtures of light and darkness. It was also believed that prisms imparted colors to light. Through observation of light refraction, Newton realized that this theory was incorrect. He demonstrated that a prism decomposes "white" light into a spectrum of colors. Newton obtained a triangular prism and began "to try therewith the celebrated Phaenomena of Colours." In his notes he states, "having darkened my chamber, and made a small hole in my window-shuts, to let in a convenient quantity of the Sun's light, I placed my Prisme at this entrance, that it might be thereby refracted to the opposite wall. It was at first a very pleasing divertisement, to view the vivid and intense colours produced thereby; but after a while applying

Newton, (Sir) Isaac,

Fig. 1 Newton's sketch of refracting sunlight with a prism. The experiment shows light passing through a hole and then collected by a lens and focused on the prism which then refracts white light into its spectral components. The experimental sketch also shows a second prism that is used to show a narrow band light is not further decomposed. A separate experiment was used to show the reversibility of the process



myself to consider them more circumspectly, I became surprised to see them in an oblong form, which, according to the received laws of Refraction, I expected should have been circular” [7]. The original sketch demonstrates the dark room environment where this experiment was conducted (Fig. 1). He was the first to use the word spectrum (Latin for “appearance” or “apparition”) in this sense. He also used additional prisms to recombine the split components and showed that a spectral colored light does not change its properties by separating out a colored beam and shining it on various objects regardless of whether it was reflected or scattered or transmitted. With respect to colors Newton said “For the rays, to speak properly, are not coloured. In them there is nothing else than a certain power and disposition to stir up a sensation of this or that Colour” [4]. During this period he also investigated the refraction of light and demonstrated that a multicolored spectrum produced by a prism could be recomposed into “white” light by a lens and a second prism [8]. In the same period Newton observed that the angle of refraction of different colors by a prism is different [9]. He noted that even when light rays in the form of a circular beam enter a prism, the spectrum of colors that exit in the position of minimum deviation is oblong. The length of the colored spectrum was in fact about five times as great as its breadth. He stated that it is the interaction of the object with light that creates the color of the object and not the object itself. This is known as Newton’s theory of color.

From this work, Newton concluded that the lens of any refracting telescope would be adversely affected by the dispersion of light into colors. This is also referred to as chromatic aberration. He constructed the first reflecting telescope in late 1668 using a mirror to avoid this problem, and this resulted in the first known functional reflecting telescope, which is today known as a Newtonian telescope [10]. Newton used a custom composition of highly reflective speculum metal to grind his own mirrors and then employed what is now known as “Newton’s rings” to judge the quality of the optics for his telescopes and demonstrated his device to the Royal Society in 1671 [11]. Their interest resulted in Newton publishing his notes, *Of Colours* [12], which was later expanded into his book *Opticks*. Newton was the first to demonstrate the use of a prism as a beam expander in his book and he describes, via diagrams, the use of multiple-prism arrays. Some 278 years after Newton’s discussion, multiple-prism beam expanders became central to the development of narrow-linewidth tunable lasers.

In 1704, Newton also discussed the corpuscular theory of light whereby light is considered to be made up of extremely small particles (which we now call photons). Newton argued that light is composed of particles or corpuscles, which were refracted by accelerating into a denser medium. To transmit forces between particles, Newton posited the existence of the ether. However, he replaced ether with occult forces based on Hermetic ideas of attraction and repulsion between

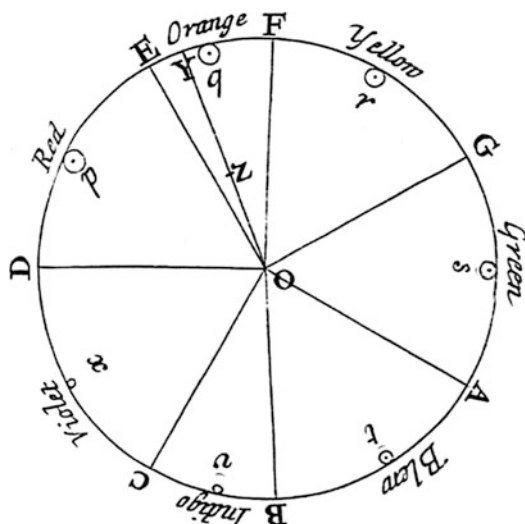
particles and Newton's considerable writings on alchemy. It has been said that "Newton was not the first of the age of reason: He was the last of the magicians." Indeed, Newton's interest in alchemy cannot be isolated from his contributions to science [11] since during his time there was no clear distinction between alchemy and science.

Later physicists favored a purely wavelike explanation of light to account for the interference patterns and the general phenomenon of diffraction. Later on, Young and Fresnel combined Newton's particle theory with the wave theory and indicated that color is the visible manifestation of light's wavelength.

The Color Circle

In his observation of white light dispersion into spectral components, Newton divided the spectrum into seven named colors: violet, indigo, blue, green, yellow, orange, and red. He placed the colors in a circular fashion and called it a color circle and used it to demonstrate the results of mixtures of spectral lights, including complementarity (those that when mixed cancel each other out and result in white, gray, or black). The size of each segment differed from the other according to his calculations of its wavelength, the seven musical tones or intervals of the eight sounds and of its corresponding width in the spectrum. The choice of seven colors was out of a belief, derived from the ancient Greek, that colors, objects in the solar system, the musical notes, and the days of the week were connected. This likely explains the selection of indigo as another hue between blue and violet. Some argue, however, that in Newton's prismatic colors, "indigo" would be placed as a color that is today called blue, whereas "blue" would correspond to cyan (Fig. 2).

Newton proposed a method to determine the "fullness or intenseness" of combined colors on the circle based on the distance of the center of the combined gravity of the circles for each of the rays of light from whiteness (the distance from O to Z in the figure shown for Color Y, which arises from the composition of all the colors in the given mixture). The color circle was, if lacking the non-spectral colors, an early representation of what became centuries later, in a modified form, the



Newton, (Sir) Isaac, Fig. 2 Colors and the associated musical notes in Newton's color wheel, shown in his book *Opticks* of 1704. The circle completes a full musical octave, from D to D [8]. Colors on opposite sides are complementary

chromaticity diagram. In his system, Newton had connected violet to red in the circle, and thus, a large gamut of purples was not shown. Newton described the complementary colors and their mixture and stated "If only two of the primary colours which in the circle are opposite to one another be mixed in an equal proportion, . . . , the colour compounded of these two shall not be perfectly white, but some faint anonymous colour." [8]. Newton's concept of complementary colors was demonstrated more thoroughly in the nineteenth century by color theorists. Helmholtz established the complementary stimulus pairs, and Ogden Rood (1831–1902) emphasized that to reveal applied colors in their natural brilliance, a knowledge of the complementary hues was required [13].

A version of Newton's color circle without the indigo blue was adopted by painters to describe complementary colors. Nonetheless, this circular diagram became the model for many color systems of the eighteenth and nineteenth centuries. The conceptual arrangement of colors in this form also allowed the painters' primaries (red, yellow, blue) to be arranged opposite their complementary colors (e.g., red opposite green), as a way of denoting that each complementary color would

Newton, (Sir) Isaac,

Fig. 3 Newton was the Warden and then the Master of the Royal Mint in England, and his picture appeared on the one pound note, almost three centuries later [15]. (With permission to reprint from the Bank of England)

**Newton, (Sir) Isaac,**

Fig. 4 An image of a German stamp commemorating Newton's 350th birthday [16]



enhance the other's effect through optical contrast.

In addition to his work on optics, Newton made seminal contributions to several other scientific disciplines. In his book the *Principia* or “*Mathematical Principles of Natural Philosophy*,” which was published in 1687, Newton formulated the laws of motion and universal gravitation and is considered to have laid the foundations for classical mechanics [14]. He also introduced the notion of a Newtonian fluid, studied the speed of sound, and developed an empirical law of cooling among other major contributions made to scientific discovery. Newton's image appeared on the Bank of England notes for about 10 years in the 1970s and 1980s [15]. He has also been commemorated on various stamps and coins [16] (Figs. 3 and 4).

References

1. Storr, A.: Isaac Newton. *Br. Med. J.* **291**, 1779 (1985). <https://doi.org/10.1136/bmj.291.6511.1779>
2. Westfall, R.S.: *Never at Rest*. Cambridge University Press (1980, 1998). ISBN 0-521-27435-4
3. Westfall, R.S.: *Isaac Newton*. Cambridge University Press (2007). ISBN 978-0-19-921355-9
4. More, L.T.: *Isaac Newton A Biography 1642–1727*. pp. 246, 381, and 389. Scribners, New York (1934)
5. Voltaire, F.M.A.: Cassell & Co. In: Price, D. (ed.) *Letters on England*. p. 100. (1894). Also available in pdf form http://livros.universia.com.br/?dl_name=Letters-on-England-de-Voltaire.pdf
6. Newton, I.: *Hydrostatics, Optics, Sound and Heat*. Cambridge University Digital Library (c. 1670–c. 1710)
7. Whittaker, E.T.: *A History of the Theories of Aether and Electricity: From the Age of Descartes to the Close of the Nineteenth Century*. Hodges, Figgis, Dublin (1910)

8. Newton, I.: *Opticks, or A Treatise of the Reflections, Refractions, Inflections & Colours of Light*. Dover Publications, New York (1952) Online version available on <https://archive.org/details/opticksoratreat00newtgoog>
9. Ball, W.W.R.: *A Short Account of the History of Mathematics*. Dover, New York (1908). ISBN 0-486-20630-0
10. Darrigol, O.: *A History of Optics from Greek Antiquity to the Nineteenth Century* (2012). ISBN-13: 978-0199644377
11. White, M.: *Isaac Newton: The Last Sorcerer*. Fourth Estate Limited (1997). ISBN 1-85702-416-8
12. Newton, I.: *Of Colours*. The Newton Project. <http://www.newtonproject.sussex.ac.uk/view/texts/normalized/NATP00004>. Retrieved 4 Apr 2015
13. Rood, O.N.: *Modern Chromatics, with Application to Art and Industry*. New York (1879). Also available online: <http://lcweb2.loc.gov/service/gdc/scd0001/2010/20100701001mo/20100701001mo.pdf>
14. Newton, I.: *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press, Berkeley (1999)
15. <http://www.royalmintmuseum.org.uk/history/people/mint-officials/isaac-newton/index.html>. Visited 9 Apr 2015
16. http://th.physik.uni-frankfurt.de/~jr/gif/stamps/stamp_newton.jpg. Visited 9 Apr 2015

Nickerson, Dorothy

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Dorothy Nickerson (August 5, 1900–April 25, 1985) was an American color scientist and

technologist who made important contributions in the fields of color quality control, technical use of colorimetry, relationship between color stimuli and color perceptions, standardization of light sources, color tolerance specification, and others. Nickerson was born and raised in Boston and attended Boston University in 1919 and Johns Hopkins University in 1923. Later, she continued her education at summer courses and university extensions at Harvard University, George Washington University, and the Graduate School of the US Department of Agriculture. Her special interest was the science of color then in significant development.

In 1921, Nickerson joined the Munsell Color Company as a laboratory assistant and secretary to A. E. O. Munsell who in 1918 had taken over the firm from his father. In 1922, the firm moved to New York and in 1923 to Baltimore. In 1927, she was offered a position at the US Department of Agriculture where she remained until her retirement in 1964. When she joined, color science and technology were without international standards and at the beginning of industrial use. Nickerson was instrumental in developing the technology and its use in agricultural and industrial settings.

Nickerson became the first individual member of the Inter-Society Color Council, founded in 1931. She was a lifelong member, received the organization's Godlove Award, and had an award named after herself. She was a member of the US National Committee of the CIE and the International Association on Color where she received the first D. B. Judd AIC Award in 1975. Nickerson was a trustee of the Munsell Color Foundation since 1942, was its president from 1973 to 1975, and assisted in the transfer of the foundation to the Rochester Institute of Technology in 1983 where it helped fund the then new Munsell Color Science Laboratory. Nickerson was the author and coauthor of some 150 articles and other publications [1].

Major Accomplishments/Contributions

Color quality control of agricultural products: In the late 1920s, Nickerson worked on usage of disk

color mixture to define the color quality of cotton and other agricultural products (see portrait image, ca. 1930, US Government) and the conversion of disk mixture data into the CIE colorimetric system [2].

Standardization of light sources for color assessment and color rendering: In the late 1930s, a major occupation of her was the development of defined light sources for visual assessment of color quality. Later, she was also active in the development and promotion of standard methods for the definition of color rendering of lights [3].

Munsell color system and its colorimetric definition: In 1940, a technical committee of the Optical Society of America (OSA) began a study of the Munsell color system, improvements and extensions of the system, and its definition in the CIE colorimetric system. Nickerson was an important participant in this effort. The final report of the committee was authored by S. M. Newhall, D. Nickerson, and D. B. Judd, and its results are known as the Munsell renotations, the specification of the aim colors of the current system. Nickerson prepared plots of the Munsell color stimuli in the CIE chromaticity diagram that remain in publication today [4, 5]. She also wrote an extended history of the Munsell system [6].

Color tolerance specification: In 1936, Nickerson published the first color difference formula for industrial use, based on the addition of increments of Munsell hue, chroma, and lightness scale values. In 1943, together with Newhall, she published realistic representations of a three-dimensional perceptually approximately uniform optimal object color solid. In 1944, together with her assistant K. F. Stultz, she published a colorimetric color difference formula, known as the Adams–Nickerson–Stultz formula, that in modified form eventually became the CIE $L^*a^*b^*$ color space and difference formula [7, 8].

Color charts: In the mid-1940s, Nickerson was active in methods for assessing the color of soils, an effort that found its expression in the Munsell Soil Color Chart, still in use today. In 1957, Munsell issued the Nickerson Color Fan, a color

fan for horticultural purposes [9]. Working with D.B. Judd, the chair of the OSA committee that developed the OSA Uniform Color Space, Nickerson, as a member of the committee, was also a contributor to that effort for over 25 years and wrote a detailed history of the development of the system [10].

References

1. Bartleson, C.J., Luke, J.T.: Dorothy Nickerson, 1900–1985. *Optics News*. **11**, 28–29 (1985)
2. Nickerson, D.: A Method for Determining the Color of Agricultural Products, U.S. Dept. of Agriculture Technical Bulletin, Washington, DC 154, 32 p. (1929)
3. Nickerson, D.: Artificial daylighting for color grading of agricultural products. *J. Opt. Soc. Am.* **29**, 1–9 (1939)
4. Newhall, S.M., Nickerson, D., Judd, D.B.: Final report of the OSA Subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Am.* **33**, 385–418 (1943)
5. Wyszecki, G.: Stiles, W.S.: *Color Science*, 2nd edn, pp. 853–861. Wiley, New York (1985)
6. (a) Nickerson, D. History of the Munsell color system and its scientific application. *J. Opt. Soc. Am.* **30**, 575–585 (1940); (b) Nickerson, D.: History of the Munsell color system. *Color Eng.* **7**, 42–51 (1969)
7. Nickerson, D.: The specification of color tolerances. *Textile Res.* **6**, 505–514 (1936)
8. Nickerson, D., Stultz, K.F.: Color tolerance specification. *J. Opt. Soc. Am.* **34**, 550–570 (1944)
9. Nickerson Color Fan, produced by the Munsell Color Company beginning in 1957, with 262 color samples in 40 hues, no longer produced
10. Nickerson, D.: History of the OSA Committee on Uniform Color Scales. *Optics News*, Winter, 8–17 (1977)

Non-photorealistic Rendering

Jorge Lopez-Moreno
GMRV Group, Universidad Rey Juan Carlos,
Madrid, Spain

Synonyms

Artistic rendering; Computer depiction; Expressive rendering

Definition

In computer graphics, non-photorealistic rendering (NPR) describes all the depiction techniques which do not aim to convey photographic realism. The reason why this term is defined by “what it is not” is that through the last decades, the computer graphics community has focused most of its efforts in creating images indistinguishable from reality (like a photograph or photorealistic), thus assimilating the subfield “photorealistic rendering” to the more general term “computer graphics.”

Some of the most well-known NPR techniques are 3D rendering with *cel shading* to simulate cartoon style, simulation of traditional art media (ink, pencil, watercolor, etc.) and artistic styles (abstract painting, impressionism, etc.), and technical depiction of objects in order to enhance clarity over realism: architecture blueprints, engineering cutaway, and explode views or anatomical dissections.

Origin of the Term

The term was coined by Salesin and Winkenbach in 1994 [1] and although its widespread use by researchers around the world, it is not free from controversy. Since the first International Symposium on Non-photorealistic Animation and Rendering (the most well-known research conference on this subject) in the year 2000 [2], there is an ongoing discussion over the looseness of this term, and several alternatives have been proposed such as “computer depiction” [3]. One reason is that *non-photorealistic* only describes a style of generating images, while the range of topics in NPR research encompasses both rendering aspects (multiple styles of picture generation) and interaction issues (e.g., artistic tools such as a simulated brush).

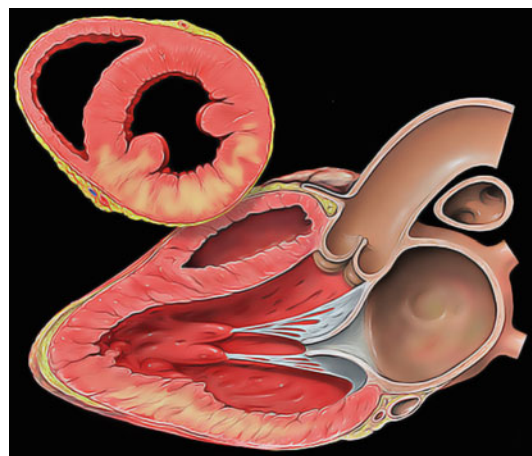
Related Fields: Art and Perception

The motivation for NPR is that realism by itself is not always desirable in order to communicate

something about the scene to be depicted. For instance, painters like Van Gogh or Munch used extremely saturated and unrealistic colors and paint strokes in order to convey not only the depicted object but also a certain mood or feeling. Likewise but not for artistic reasons, stylized illustration is preferred over real images in medicine to teach anatomy, as they help in differentiating the parts of the human body which are otherwise difficult to identify in their natural state (see Fig. 1).

Given this range of applications, the NPR field has naturally embraced interdisciplinary collaboration both with the artistic community and human perception researchers: in the last decade as applications like augmented reality move away from photorealism in a wish for conveying additional or specific information (e.g., overlaid pathology data during a surgery intervention), a strong need arises for checking that the depiction that they do provide faithfully communicates the desired information.

This interaction between art, perception, and NPR is widespread in the literature. Researchers have studied how humans paint or draw in order to derive models which predict the success of NPR algorithms [4]. For instance, by combining several depictions of the same object made by different persons, it is known that most of the strokes



Non-photorealistic Rendering, Fig. 1 Illustration of a heart inferior wall dysfunction (short and long axis views). (By Patrick J. Lynch (medical illustrator) and C. Carl Jaffe (MD, cardiologist))

can be predicted thanks to three-dimensional features such as contours or curvature and thus are automated by line drawing methods like *suggestive contours* [5]. However, it is also evident in these studies that even the most recent and sophisticated methods fail to predict all the elements used by artists to draw an object.

Artistic styles have also been analyzed and, when possible, simulated from the physical media (e.g., simulation of fluids for paint) to the gamut of colors or the strokes and the shapes [6]. Rendering in a specific style usually implies a combination of multiple NPR techniques: Fig. 2 shows some of the image processing methods used for rendering stylized depictions. Although most of these techniques still require a significant amount of user input (in choosing and combining multiple parameters), in recent years, some approaches for automatic abstraction or color and histogram matching have been proposed [7, 8].

In recent years, algorithms based on deep learning have achieved very good results of painterly style transfer, that is, transforming real photographs or synthetic renderings into exemplars of specific artistic styles. This is the case of the work

by Luan et al. [9] where they achieve multiscale harmonization of photographs inserted into paintings by means of a VGG network and two consecutive passes for coarse harmonization and subsequent localized detail texture similarity enforcement.

Classification

The works in the field are usually classified by their input system (2D, 3D) and type of user input (assisted or automatic):

- **From 3D to 2D:** Given a 3D geometric description of the objects, their material properties and light sources, a 2D image is generated according to a non-photorealistic style. For instance, in the cel shading technique, hand-drawn cartoon style is mimicked by adding a black contour line (occluding contour) based in the culling of front faces in the geometric model and illuminating the object with a flat shading obtained by limiting the range of colors to 2 or 3 t. The continuous value yielded by traditional shading formulae

N

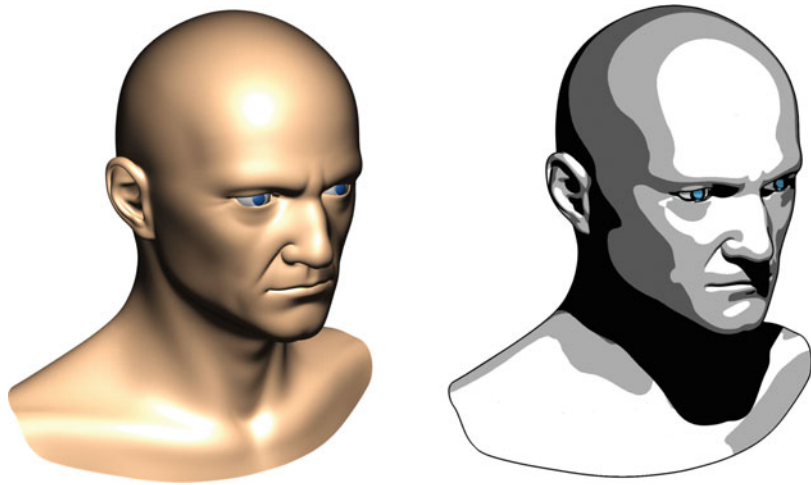


Non-photorealistic Rendering, Fig. 2 Some examples of 2D NPR stylization. *Top row:* input images. *Bottom row:* resulting depictions, from *left to right* – oil strokes

simulation based on flow integration (LIC), edge detection with color posterization, and abstraction based in hue shift and pixel splattering

Non-photorealistic

Rendering, Fig. 3 Toon shading render generated with available commercial software. The 3D model in the *left* was created with a subsurface scattering shading technique. The NPR version in the *right* includes shadows, contour lines, and three shades of gray



(based on the cross product of the vertex normals and the direction of the light source) is thresholded to a limited set of values to produce this effect (Fig. 3). Nowadays, most of these rendering effects are generated as a postprocessing step in image space. Surface normals, depth, and reflectance values are computed per pixel and stored as images in different layers (a technique known as deferred shading). In this context, image processing is used to the same effect. For instance, contour lines can be generated by detecting discontinuities in the normal image with an edge filter such as Canny's detector [10].

- **From 2D to 2D:** This is usually based on image processing techniques and no user intervention. Most of the commercially available photo cameras provide at least a few of these transformations. Given a photograph as input, a purely 2D cel shading effect is obtainable by a combination of histogram thresholding (*posterize*) for color and a contour detector (Canny) for the contour lines. The simulation of brush and pencil strokes by using *line integral convolution* (LIC) [11] is another well-known example, which integrates random sampling from the luminance vector field of the original image in order to generate lines emulating the most likely strokes done by artists (see bottom left image in Fig. 2). Saliency

and feature detection are also considered to emphasize those strokes which have greater visual significance.

- **From 2.5D to 2D:** That is, 2D plus additional data such as incomplete 3D information (depth range or *Z-buffer*) and/or user annotations – semantic tags (*this part of the image is a cloud*), strokes (which encode information in its shape), marks (*this part of the object is on top to the other*), etc. Some recent techniques [12] include perception assumptions (such as *global convexity* and *saliency*) in order to infer depth from 2D images. This depth can further be edited (sculpted) with brush strokes or directly used in combination with traditional 3D-to-2D algorithms to render NPR images like those in Fig. 4.
- A fourth possibility is **from 2D to 3D**, that is, from a set of user strokes or images (or a combination of both) of a given scene, a 3D model is derived which allows for novel views. This conforms a field by itself called *image-based modeling and rendering* (IBMR), which lies between the areas of computer graphics and computer vision.

Examples

Since the early adoption of cel shading in the title *Jet Set Radio* (year 2002), hundreds of examples



Non-photorealistic Rendering, Fig. 4 Example of perception-based NPR rendering from Ref. [12]. *Top left*: original photo. *Top right*: multitone depiction obtained

from hallucinated depth and virtual light sources. *Bottom left*: additional comic book style result with two layers of dynamic lines added. *Bottom right*: NPR color relighting

N

of NPR techniques have emerged in video games. Likewise, NPR is ubiquitous in animation and films. The *South Park* series is an unexpected example of 2D cartoon style derived from 3D models. A case of 2D real-life footage to 2D hand-drawn animation style is shown in the film *A Scanner Darkly*, where key frames were roto-scoped and vectorized, automatically interpolating the frames in between. However, although some automation processes were introduced, this production required a significant amount of user input.

More recently, the aesthetic currents reminiscent of 8-bit and 16-bit games and systems have contributed to the popularity of *pixel art*, where images and animations (sprites) are edited on the pixel level, with deliberate control over the placement of individual pixels, often with the intent of maximizing the expressivity of the rendering with a minimalistic set of lines, colors, and gradients



Non-photorealistic Rendering, Fig. 5 Example of Pixel Art, artwork derived from videogame *Sancho Island*

(see Fig. 5). *Pixel art* is usually rendered in 2D, but upcoming games (year 2021) such as *The last Night* or *Replaced* have translated this technique to hybrid 3D environments, combining it with physically based illumination and complex volumetric shading effects.

Cross-References

- [Color Harmony](#)
- [Compositing and Chroma Keying](#)
- [Global Illumination](#)
- [Optical Art](#)
- [Sampling Problems in Computer Graphics](#)

References

1. Winkenbach, G., Salesin, D.H.: Computer-generated pen-and-ink illustration. In: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH, New Orleans, Louisiana, USA, pp. 91–100 (2008)
2. Fekete, J.-D., Salesin, D. (eds.): Proceedings of the 1st International Symposium on Non-photorealistic Animation and Rendering. Annecy, France. ACM (2000)
3. Durand, F.: An invitation to discuss computer depiction. In: Proceedings of the 2nd International Symposium on Non-photorealistic Animation and Rendering, Annecy, France, pp. 111–124. ACM (2002)
4. Cole, F., Golovinskiy, A., Limpaecher, A., Barros, H.S., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S.: Where do people draw lines? Commun. ACM. **55**, 107–115 (2012)
5. DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., Santella, A.: Suggestive contours for conveying shape. ACM Trans. Graph. **22**(3), 848–855 (2003)
6. Gooch, B., Gooch, A.: Non-photorealistic Rendering. A K Peters, Natick (2001)
7. Zhao, M., Zhu, S.: Sisley the abstract painter. In: Proceedings of the 8th International Symposium on Non-photorealistic Animation and Rendering Annecy, France, pp. 111–124. ACM (2010)
8. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2001)
9. Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep Painterly Harmonization. In: Eurographics Symposium on Rendering EGSR (2021)
10. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
11. Cabral, B., Leedom, L.: Imaging vector fields using line integral convolution. In: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '93, Anaheim, pp. 263–270 (1993)
12. Lopez-Moreno, J., Jimenez, J., Hadap, S., Anjyo, K., Reinhard, E., Gutierrez, D.: Non-photorealistic, depth-based image editing. Extended papers from NPAR 2010. Comput. Graph. **35**(1), 99–111 (2011)

Nonvisual Lighting Effects and Their Impact on Health and Well-Being

Mariana Figueiro

Lighting Research Center, Troy, NY, USA

Synonyms

[Circadian rhythms](#); [Circadian system](#)

Definition

Biological, nonvisual effects of lighting constitute a growing field within lighting research and education. In addition to visual effects, light is known to affect other biological rhythms, most notably the circadian system, which generates and regulates a number of rhythms that run with a period close to 24 h. Light/dark patterns incident on the retina are the major synchronizer of circadian rhythms to the 24-h solar day. Lighting characteristics (quantity, spectrum, timing, duration, and distribution) affecting the visual and circadian systems differ. Symptoms of circadian sleep disorders, such as seasonal affective disorder, jet lag, and delayed sleep phase disorder, can be mitigated by timed light exposure.

Introduction

The neurophysiology and neuroanatomy of the human visual system are largely understood: vision is the result of complex interactions between light sources, objects and surfaces, the eye, and the brain. Lighting engineers and practitioners have developed technologies, standards,

measurement devices, and applications based on this knowledge, and one of the primary goals of lighting design is to increase performance and productivity by improving vision and perception. Recent scientific discoveries, however, have added another dimension to lighting: namely, that *light is not just for vision anymore*. Specifically, scientists have discovered that daily light/dark patterns reaching the retina (back of the eye) play an important role in human health and well-being because they regulate our bodies' circadian rhythms. Circadian rhythms are every rhythm in our body that repeats approximately every 24 h. The circadian system is profoundly important for many human behaviors, as well as human well-being.

This entry provides basic knowledge about the nonvisual effects of light entering the eye, focusing exclusively on the effects of light on human circadian and diurnal rhythms. Also discussed here are the lighting characteristics that affect these rhythms, as well as applications in which light/dark cycles have been experimentally shown to affect them.

Overview of the Circadian System

Earth rotates around its axis, and, as a result, all creatures exposed to daylight on Earth experience a 24-h cycle of light and dark. Living organisms have adapted to the planet's daily rotation by developing biological rhythms that repeat at approximately 24-h intervals [1]. These rhythms are called circadian rhythms, from the Latin *circa* (about) and *dies* (day).

Biological Clock

Circadian rhythms are generated internally by the body, yet they are constantly aligned with the environment by factors that are external to the body, mainly light/dark cycles. In mammals, circadian rhythms are regulated by an internal biological clock located in the suprachiasmatic nuclei (SCN) of the hypothalamus of the brain. The biological clock in humans has a natural period that is slightly greater than 24 h, and

environmental cues, such as light/dark cycles, social activities, and meal times, can reset and synchronize the clock daily, ensuring that our behavioral and physiological rhythms are synchronized with the external environment.

Output Rhythms

Output rhythms are behavioral and physiological rhythms regulated by the biological clock, such as the sleep/wake cycle, alertness, core body temperature, locomotor activity, feeding and drinking behavior, and hormone production. Of special interest here are the sleep/wake cycle, alertness, core body temperature, and hormone production, especially melatonin production.

Sleep/Wake Cycle and Alertness

The sleep/wake cycle is one of the most prominent circadian rhythms. We are under the influence of two opposing systems: the sleep drive (homeostatic) and the alerting force (circadian). The sleep drive and the alerting force are distinct forces, independent from each other, though complementary, ensuring that we are asleep at night and awake during the day. The sleep drive, for example, is low when we get up, increasing steadily throughout the waking day, before diminishing rapidly within the first hours of sleep. The alerting force is regulated by the biological clock and follows a circadian rhythm, reaching a peak during the early evening and a trough during the second half of the night. The interaction between the sleep drive and the alerting force determines when we fall asleep and how well we sleep at night. Alertness is strongly correlated with the sleep/wake cycle.

Core Body Temperature

Core body temperature also follows a circadian pattern. It is high during the day, reaching a peak in the early evening, and low at night, reaching its low point about 2 h before one naturally wakes up. Core body temperature is typically in phase with alertness and has a negative correlation with the hormone melatonin. The alerting force is at its minimum during the second half of the night, close to the core body temperature minima.

Hormone Production

Although there is evidence that the biological clock influences a variety of hormones, this discussion will be limited to melatonin production by the pineal gland. The pineal gland is located near the center of the brain, is about the size of a pea, and has a shape resembling that of a pinecone. It is believed that the primary function of the pineal is to convey light/dark information to the body via secretions of the hormone melatonin [2]. Melatonin is easily absorbed into the bloodstream, which makes it an ideal chemical messenger of time of day information for the entire body. Melatonin also participates in the transmission of information concerning day length or photoperiod for the organization of seasonal responses in animals (e.g., breeding).

Melatonin is produced during the subjective night (i.e., inactive period in diurnal species and active period in nocturnal species) and under conditions of darkness. As a result, melatonin levels are low during the day and high at night (in the dark). Although not yet well established, many researchers believe that melatonin can induce sleep in diurnal mammals, including humans, by acting on the SCN and reducing the alerting force, or the “wake-promoting” signal sent by the biological clock during the daytime. In addition, melatonin has an inverse relationship with core body temperature, which is closely linked to waking and sleep times. Peak melatonin levels typically occur slightly before core body temperature bottoms out. When external cues are absent, melatonin starts being endogenously produced by the body at around 9–10 p.m. (or about 2 h prior to sleeping) and stops being produced at around 7–9 a.m. Maximum melatonin levels occur between 2 and 4 a.m [2].

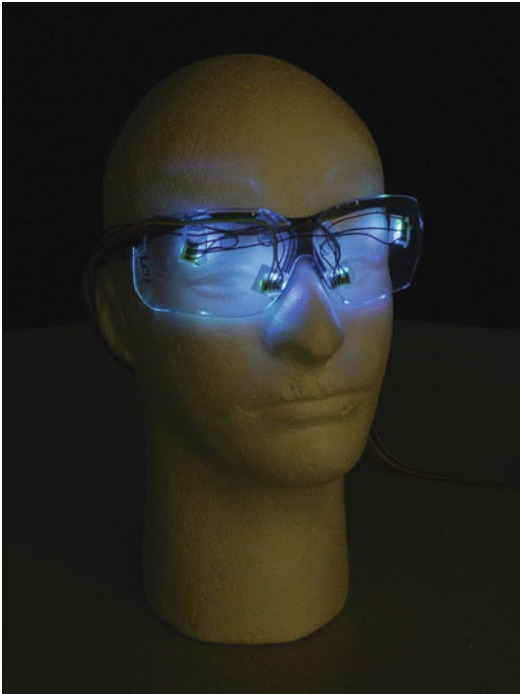
Regulation of Circadian Rhythms by Light

The 24-h light/dark cycle is the biological clock’s main synchronizer to the solar day. Light can phase advance or phase delay human circadian rhythms, depending upon when it is applied [3]. For example, light that is applied before the

minimum core body temperature, which is reached approximately 1.5–2.5 h before we naturally awaken, will delay the clock (e.g., one will wake up later the following day), and light applied after minimum core body temperature is reached will advance the clock (e.g., one will wake up earlier the following day). Although light is the main synchronizer of the biological clock to the solar day, it is not the only one. Exercise, social activities, timing of the sleep/wake cycle, and scheduled meals have also been shown to shift and synchronize the clock, although their impact on circadian rhythms seems to be weaker than the impact of light/dark cycles.

Lighting Characteristics Affecting Circadian Rhythms

The neural machinery in the mammalian retina provides light information to both the visual and circadian systems, but the two systems process optical radiation (light) differently [3]. Rods, cones, and the intrinsically photosensitive retinal ganglion cells (ipRGCs) [4] participate in circadian phototransduction (how the retina converts light signals into electrical signals for the biological clock). The quantity of polychromatic “white” light necessary to activate the circadian system is at least two orders of magnitude greater than the amount that activates the visual system. The circadian system is maximally sensitive to short-wavelength (“blue”) light, with a peak spectral sensitivity at around 460 nm (Fig. 1), while the visual system (i.e., acuity) is most sensitive to the middle-wavelength portion of the visible spectrum, with a peak at around 555 nm. Operation of the visual system does not depend significantly on the timing of light exposure and, thus, responds well to a light stimulus at any time of the day or night. On the other hand, the circadian system is dependent on the timing of light exposure, as discussed above. In addition, while the visual system responds to a light stimulus very quickly (in milliseconds), the duration of light exposure needed to affect the circadian system can take minutes. For the visual system, spatial light distribution is critical for good visibility. It is not yet well established how light incident on different portions of the retina affects the circadian system.



Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 1 Photo of light goggles engineered by the Lighting Research Center. For the same photopic lux at the eye, “blue,” short-wavelength light has the greatest impact on the circadian system

It is also important to note that the short-term history of light exposure affects the sensitivity of the circadian system to light; the higher the exposure to light during the day, the lower the sensitivity of the circadian system to light, as measured by nocturnal melatonin suppression and phase shifting.

Specifying Light for the Circadian System

In 2005, Rea et al. proposed a mathematical model of human circadian phototransduction based on the neuroanatomy and neurophysiology of the human retina and using published action spectrum data for acute melatonin suppression [5, 6]. Using this model, spectral irradiance at the cornea is first converted into circadian light (CL_A), which is comparable to photopic illuminance but weighted by the spectral sensitivity of the human circadian system (Fig. 2a) as measured by acute melatonin suppression after a 1-h exposure. Then the circadian stimulus (CS) is determined, which reflects

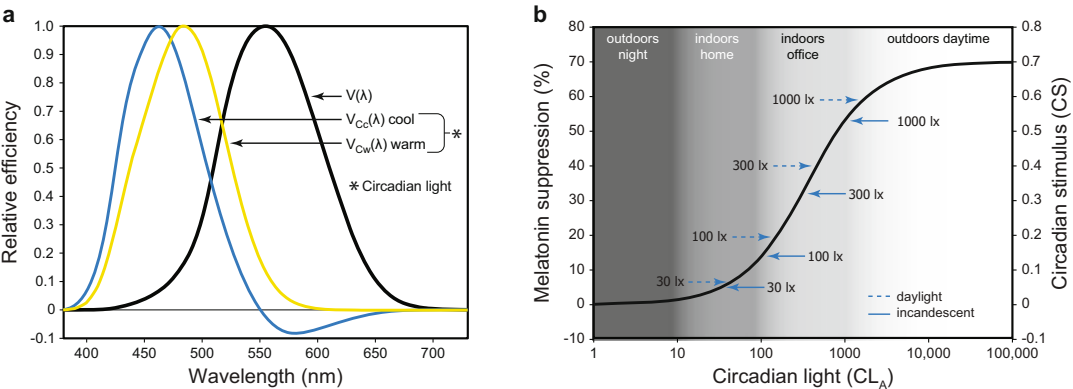
the effectiveness of the spectrally weighted irradiance at the cornea from threshold ($CS \approx 0.1$) to saturation ($CS \approx 0.7$) (see Fig. 2b). While CS was developed using data from studies that measured acute melatonin suppression, Zeitzer et al. [7] showed that acute melatonin suppression and phase shifting of melatonin rhythms followed similar threshold and saturation response characteristics. The model has been successfully used to predict the effectiveness of various light sources for activating the circadian system in laboratory [8–13] and field studies [14–18].

Measuring Light for the Circadian System

Given that lighting characteristics affecting the visual system are different from those affecting the circadian system and that commercial light meters are calibrated to measure light for the visual system, new measurement devices that properly characterize light for the circadian system are needed. Two similar devices, the Daysimeter-S and the Daysimeter-D (also known as Dimesimeter), have been developed [19]. They are personal circadian light meters that measure light that is effective for visual and circadian responses. The Daysimeter-S has the photosensor package positioned near the plane of one cornea, and the Daysimeter-D can be worn as a pin, as a pendant, or on the wrist (Fig. 3). These photosensors are calibrated to measure “circadian” light. They also provide accurate measurements of light for vision (photopic lux). The photosensor package closely matches the eye’s acceptance of angles of light. Since circadian rhythms follow a 24-h pattern, it is necessary to measure light (and dark) over extended time periods to ascertain the stability and period of circadian rhythms. To this end, the Daysimeters can be deployed in the field to gather light data for up to 30 days with a 30-s sampling interval.

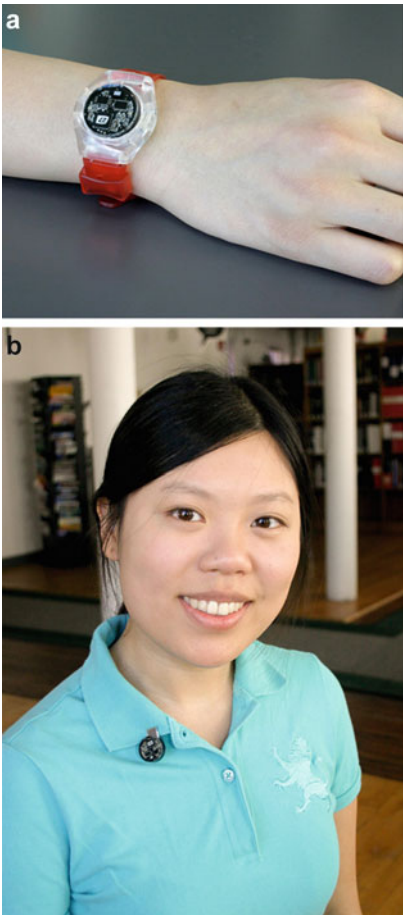
Applications

In the following section, practical situations in which the effects of light on the circadian system could be beneficial to human health and well-being will be discussed. It is important to note



Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 2 The photopic illuminance ($V(\lambda)$) circadian light (vision) spectral efficiency

functions (a) and the absolute sensitivity of the human circadian system to light (b)



Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 3 (a, b) Photo of the Daysimeter-D worn on the wrist and as a pin

that light is not currently known to cure any diseases. Further, individual differences need to be considered when designing lighting for the circadian system, so the lighting recommendations discussed below should be used as a framework only, and adjustments must be considered. It is also important to emphasize that few of the studies using light as a non-pharmacological treatment for various diseases and disorders reported the spectral power distribution of the light source, which makes it difficult to quantify the impact of light on the retina for circadian effectiveness. Consequently, generalizations from these studies are often more qualitative (e.g., bright vs. dim) than quantitative.

Tentative application tips will be made, and the quantity, spectrum, duration, and timing of the light presented at the eye will be discussed. In order to simplify and standardize the tips offered here, **bright white light** will be defined as 1-h exposure to a CS of 0.3 (at least 500 lx measured at eye level from a white light source), unless otherwise stated in the text. This light level is about five times the amount of light one typically finds in an office environment without a window. “Blue” light is defined here as 1-h exposure to 40 lx at the cornea from a narrowband light source peaking at 470 nm. If duration of exposure is increased, light levels may be decreased, but it is recommended that it is never below a CS of 0.2 (roughly 300 lx at the cornea for the white light)

and 20 lx at the cornea for the blue light. The lighting recommendations offered here are based on current knowledge from research findings, but results of future research will refine these values.

Seasonal Affective Disorder

Seasonal affective disorder (SAD) is a subtype of depression, with episodes occurring during winter months and remitting during summer months. Symptoms of SAD include depression, hypersomnia, and weight gain due to increased carbohydrate cravings, social withdrawal, and even suicidal thoughts. It is believed that because daylight availability decreases in the winter at high latitudes, the number of people experiencing SAD increases as the latitude increases. “The winter blues” is an even more common subtype of SAD. The mechanisms of SAD are still unknown, and there are several competing hypotheses as to what causes SAD and how light can be used as a treatment. One of these is that late daybreak during winter months delays the circadian rhythms of those more susceptible to SAD; in this case, morning light is believed to be effective in treating symptoms of SAD. Another hypothesis is that the overall melatonin production of those suffering from SAD is greater during winter months than during summer months, which extends the amount of time during the 24-h day that their bodies think it is nighttime. In this case, light in the early morning or evening is recommended.

If a person is formally diagnosed with SAD by a general practitioner, insurance companies may pay for the cost of light treatment devices. A recent study [20] showed that approximately 500 lx of blue light ($\lambda_{\max} = 470$ nm) directed at the eye was able to significantly improve SAD symptoms compared to red light used for placebo control. It has been suggested, however, that the positive impact of light on SAD symptoms is simply a result of placebo effects.

Practical Tips

- Encourage SAD patients to go for a half-hour walk outdoors in the morning (right after daybreak).
- Expose individuals to bright white light (or “blue” light) in the morning or evening.

Jet Lag

Jet lag is a temporary desynchronization between the biological clock time and the environmental time (light/dark). The symptoms include insomnia and/or hypersomnia, fatigue, poor performance, and gastrointestinal problems. Eastward travel generally results in difficulty falling asleep, and westward travel results in difficulty maintaining sleep. Adaptation to a new time zone is usually slower after eastward travel than after westward travel. This is because [1] those traveling east need to advance their biological clock to readjust to local time at their destination; the time that daylight is available upon arrival at the final destination will promote phase delay of the biological clock, and [2] it is easier for the timing of the biological clock to be delayed than advanced. One study [21] showed that a combination of advancing sleep schedules for 1 h per day plus morning light treatment (one half hour of 5,000 lx + half hour of less than 60 lx at the cornea) for 3.5 h advanced the phase of the biological clock by 1.5–1.9 h in 3 days. Although the principles for applying light treatment for reducing jet lag symptoms are known, the implementation of the light treatment may be a challenge. Airlines are starting to use colored light inside airplanes to improve mood, but it is probably very difficult to shift the biological clock to promote complete adjustment to a new time zone while one is inside the plane. Because the circadian clock is slow to shift, users need to start treatment a few days before they are scheduled to travel. However, because travelers have busy schedules, the likelihood of compliance is low. A personal light treatment device could be developed to increase the likelihood of compliance (Fig. 4).

Practical Tips

- Eastward travel: Upon arrival at your destination, avoid bright white light (including daylight) during the morning, until about noon local time, and seek exposure to bright white (or “blue”) light in the afternoon. Upon arrival, wear glasses that filter out radiation below 520 nm (“orange” glasses; Fig. 5) to avoid circadian-effective light because the circadian

Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 4

Photo of a possible light box design that could aid in the adaptation of travelers' circadian systems to new time zones



Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 5

Orange goggles are able to block short-wavelength light that maximally activates the circadian system. These goggles should be worn at times when one wants to reduce circadian effective light exposures



system is maximally sensitive to short-wavelength (“blue”) light.

- Westward travel: Upon arrival at your destination, seek exposure to bright white light (or “blue” light) during the daylight hours and avoid bright white light in the evening.

Delayed Sleep Phase Disorder

Delayed sleep phase disorder (DSPD) is a disorder of sleep timing; people suffering from DSPD typically go to bed late and wake up late (3–6 h later than typical sleeping hours). This pattern interferes with people’s normal functioning because they have difficulty waking up in the

morning for work, school, and social obligations, and since they go to bed late, they do not sleep for as many hours as those going to bed at more normal hours. DSPD in adolescents is common and probably associated with hormonal changes that occur at puberty. The exact causes of DSPD are not actually known, but light exposure after minimum core body temperature and dim light during the evening have been shown to advance the phase of the biological clock of persons with DSPD [22]. Using this information, field studies [23] were conducted to investigate the impact of light exposures on dim light melatonin onset (DLMO), a primary marker for the timing of the

biological clock, and on sleep duration for two populations of eighth graders. It was hypothesized for one study conducted in North Carolina that the lack of morning short-wavelength light (which was removed by wearing orange goggles; Fig. 5) would delay the timing of the students' biological clocks. For the other study conducted in New York, it was hypothesized that exposure to more evening light in spring relative to winter would also delay the biological clocks of adolescents. In both studies, as expected, the students exhibited delayed DLMO as a result of removing short-wavelength morning light and as a result of seasonal changes in evening daylight. Also as expected, both sets of adolescents exhibited shorter sleep times; because of the delay in the timing of the biological clock, they fell asleep later but still had to get up at a fixed time in the morning. These two field studies clearly demonstrate that by controlling circadian light exposures, it is possible to practically and effectively control circadian time and thereby affect meaningful outcomes like sleep duration.

Practical Tips

- Expose individuals to bright white (or "blue") light in the morning, after the minimum core body temperature. Note that minimum core body temperature of people with DSPD may occur as late as 9 a.m.; therefore, an understanding of the characteristics of the biological clock of those using the light treatment is needed before recommending the timing of the light exposure. Just as a tip, minimum core body temperature generally occurs 1.5–2.5 h prior to waking without an alarm clock.
- Wear glasses that filter out radiation below 520 nm ("orange" goggles; Fig. 5) in the evening and in the morning, before minimum core body temperature is reached. Because the circadian system is maximally sensitive to short-wavelength light, elimination of optical radiation below 520 nm at these specific times will reduce the chances of phase shifting the biological clock to a later time, aggravating symptoms of DSPS even further.
- Avoid the use of very bright self-luminous electronic devices (e.g., computer screens, tablets, and cell phones) during the evening hours. At a minimum, dim them down or filter their screens with "orange" filters.

Sleep in Older Adults

Sleep disturbances in older adults. Seniors living in controlled environments (assisted living and nursing homes) are perhaps the best example of a population at risk for circadian sleep disorders; due to age-dependent reduced retinal light exposures and to fixed lighting conditions in their living environments, seniors are less likely to experience the necessary, robust 24-h, light/dark pattern needed for circadian entrainment. Prescribed 24-h light/dark patterns have been shown to influence circadian rhythms and thereby alleviate some sleep and agitation issues common among seniors, including those with Alzheimer's disease (AD). A 24-h lighting scheme has been proposed that delivers high circadian stimulation during the daytime hours, low circadian stimulation in the evening hours, and night lights that provide perceptual cues to decrease the risk of falls at night [24], such as those shown in Fig. 6. Studies showed that exposure to bright white light (ranging from 2,500 lx to as high as 8,000 lx at the cornea) for at least 1 h in the morning for a period of at least 2 weeks was found to improve or consolidate nighttime sleep of AD patients. Greater sleep efficiency at night decreased the need to sleep during daytime hours and, in some cases, reduced agitated behavior such as pacing, aggressiveness, and speaking loudly. Research demonstrated that exposing 22 AD patients to continuous, bright, and indirect white light (average of 1,136 lx at the eye) over 4 weeks consolidated their rest/activity rhythms [25]. In a long-term study, it has been demonstrated that light attenuated cognitive deterioration by 5% on the mini-mental state examination. Light also ameliorated depressive symptoms by 19% on the Cornell Scale for Depression in Dementia and attenuated the increase in functional limitations over time by 53% on the nurse-informant activities of daily living scale or a relative 53% difference [26].

Nonvisual Lighting Effects and Their Impact on Health and Well-Being, Fig. 6 Novel nightlighting system using horizontal/vertical cues that can improve postural control and stability



Practical Tips

- For older adults who want to fall asleep later in the evening, expose them to bright white (or “blue”) light in the late afternoon/early evening (before 7 p.m.), before minimum core body temperature. It is recommended that the duration of the exposure be increased to 2 h or the quantity of light be increased by a factor of two because of the optical changes to the aging eye.
- Wear glasses that filter out radiation below 520 nm (“orange” goggles) in the very early morning hours (before 7–8 a.m.).
- For AD patients who do not have a consistent sleep/wake schedule, expose them to bright white (or “blue”) light at any time during the day and reduce evening light exposures. Be consistent with the timing of the light exposure, however. Because the circadian phase of AD patients is not known and because they tend to remain in continuous dim light all day, it is expected that a robust light/dark pattern will positively impact their sleep.
- Install safe nightlighting systems that provide low-level illumination and perceptual (horizontal/vertical) cues to aid in controlling postural control and stability (Fig. 6).

Entrainment for the General Population

As shown in the examples above, lighting can help reduce or mitigate symptoms associated with various circadian sleep disorders. Lighting can also aid in maintaining entrainment in the healthy, general population. Because humans have a biological clock that runs with a period slightly greater than 24 h, we need daily morning light exposure to maintain entrainment to the solar day. In winter months, some of us can go to and come back from work or school in darkness; therefore lighting in our work environments should provide enough circadian stimulation during at least the morning hours. Although the link between circadian entrainment and productivity is yet to be established in larger, clinical studies, one study [27] to date showed that compared with a 4,000 K light source, blue-enriched white light (17,000 K) improved subjective measures of alertness, positive mood, performance, evening fatigue, irritability, concentration, and eye discomfort. Daytime sleepiness was reduced, and the quality of subjective nocturnal sleep was improved under blue-enriched white light. More recently, two field studies conducted in nine US federal government office buildings showed that both morning [17] and all-day [28] exposures to a

CS of at least 0.3 improved objective measures of circadian entrainment and sleep quality (actigraphy) and subjective measures of sleep disturbances, daytime sleepiness, depression, vitality, and energy (questionnaires) in office workers. Although future studies should confirm and extend these results, it is suggested that a dynamic lighting system that delivers bright white light (or “blue” light) to office workers during the morning hours can help maintain entrainment and possibly improve well-being, alertness, and performance, especially in winter months, when duration of daylight availability is short [29].

The Future

This overview of the impact of light on the circadian system and its effects on our health and well-being underscores the importance of developing a new framework for lighting practices that includes not only those lighting characteristics that affect the visual system but also those that affect the circadian system. Because there are great differences between the two systems’ responses to the quantity of light, its spectral composition, spatial distribution, timing, and duration, generalizations about “quality lighting” will have to be assessed by two very different sets of criteria in the future. Although the information and recommendations presented here will certainly be refined as more research is undertaken, little progress will be made in delivering “healthy lighting” to society until researchers and practitioners begin to consider, measure, calculate, and control the fundamental characteristics of light for the circadian system. Hopefully, this entry will serve as an important step toward that goal.

Finally, it is important to remember that we have no practical way to intuitively know when and what type of light we need for circadian entrainment. Like other medical monitors where we have no conscious access to what we need for good health (e.g., glucose monitors for diabetes), we need technologies to measure and track the state of our circadian system. A wireless technology connecting the Daysimeter to a smartphone is currently under development. A smartphone

(or similar technology) will be able to interpret the light/dark patterns to which a person is exposed and then be able to recommend, based on models of human circadian entrainment [30], both current and future light exposure patterns for the user to maintain circadian entrainment. One day, the smartphone might be able to interface with the building lighting system itself so that children in a school, nurses in a hospital, or people at home would be exposed to the proper lighting conditions for maintaining well-being and for increasing productivity.

Cross-References

- [Circadian Rhythms](#)
- [Light-Emitting Diode, LED](#)

Acknowledgments The author would like to acknowledge the projects’ sponsors (National Institute on Aging, National Institute of Nursing Research, National Institute on Drug Abuse, National Cancer Institute, and Office of Naval Research). Mark Rea, PhD, of the Lighting Research Center is acknowledged for his technical assistance, and David Pedler of Rensselaer Polytechnic Institute is acknowledged for his editorial assistance.

References

1. Moore, R.Y.: Circadian rhythms: basic neurobiology and clinical applications. *Annu. Rev. Med.* **48**, 253–266 (1997)
2. Arendt, J.: Melatonin and the Mammalian Pineal Gland. Chapman & Hall, London (1995)
3. Rea, M.S., Figueiro, M.G., Bullough, J.D.: Circadian photobiology: an emerging framework for lighting practice and research. *Light. Res. Technol.* **34**, 177–187 (2002)
4. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science*. **295**, 1070–1073 (2002)
5. Rea, M.S., Figueiro, M.G., Bullough, J.D., Bierman, A.: A model of phototransduction by the human circadian system. *Brain Res. Rev.* **50**, 213–228 (2005)
6. Rea, M.S., Figueiro, M.G., Bierman, A., Hamner, R.: Modelling the spectral sensitivity of the human circadian system. *Light. Res. Technol.* **44**, 386–396 (2012)
7. Zeitzer, J.M., Dijk, D.J., Kronauer, R.E., Brown, E.N., Czeisler, C.A.: Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *J. Physiol.* **526**, 695–702 (2000)

8. Wood, B., Rea, M.S., Plitnick, B., Figueiro, M.G.: Light level and duration of exposure determine the impact of self-luminous tablets on melatonin suppression. *Appl. Ergon.* **44**, 237–240 (2013)
9. Figueiro, M.G., Wood, B., Plitnick, B., Rea, M.S.: The impact of light from computer monitors on melatonin levels in college students. *Neuro Endocrinol. Lett.* **32**, 158–163 (2011)
10. Rea, M.S., Figueiro, M.G.: A working threshold for acute nocturnal melatonin suppression from “white” light sources used in architectural applications. *J. Carcinog. Mutagen.* **4**, 1000150 (2013)
11. Figueiro, M.G., Bierman, A., Rea, M.S.: Retinal mechanisms determine the subadditive response to polychromatic light by the human circadian system. *Neurosci. Lett.* **438**, 242–245 (2008)
12. Figueiro, M.G., Lesniak, N.Z., Rea, M.S.: Implications of controlled short-wavelength light exposure for sleep in older adults. *BMC. Res. Notes.* **4**, 334 (2011)
13. Figueiro, M.G., Bierman, A., Plitnick, B., Rea, M.S.: Preliminary evidence that both blue and red light can induce alertness at night. *BMC Neurosci.* **10**, 105 (2009)
14. Young, C.R., et al.: At-sea trial of 24-h-based submarine watchstanding schedules with high and low correlated color temperature light source. *J. Biol. Rhythm.* **30**, 144–154 (2015)
15. Figueiro, M., Plitnick, B., Rea, M.: Research note: a self-luminous light table for persons with Alzheimer’s disease. *Light. Res. Technol.* **48**, 253–259 (2016)
16. Figueiro, M.G., et al.: Tailored lighting intervention improves measures of sleep, depression, and agitation in persons with Alzheimer’s disease and related dementia living in long-term care facilities. *Clin. Interv. Aging.* **9**, 1527–1537 (2014)
17. Figueiro, M.G., et al.: The impact of daytime light exposures on sleep and mood in office workers. *Sleep Health.* **3**, 204–215 (2017)
18. Figueiro, M.G., Rea, M.S.: Office lighting and personal light exposures in two seasons: impact on sleep and mood. *Light. Res. Technol.* **48**, 352–364 (2016)
19. Figueiro, M.G., Hammer, R., Bierman, A., Rea, M.S.: Comparisons of three practical field devices used to measure personal light exposures and activity levels. *Light. Res. Technol.* **45**, 421–434 (2013)
20. Glickman, G., Byrne, B., Pineda, C., Hauck, W.W., Brainard, G.C.: Light therapy for seasonal affective disorder with blue narrow-band light-emitting diodes (LEDs). *Biol. Psychiatry.* **59**, 502–507 (2006)
21. Eastman, C.I., Gazda, C.J., Burgess, H.J., Crowley, S. J., Fogg, L.F.: Advancing circadian rhythms before eastward flight: a strategy to prevent or reduce jet lag. *Sleep.* **28**, 33–44 (2005)
22. Rosenthal, N.E., Joseph-Vanderpool, J.R., Levendosky, A.A., Johnston, S.H., Allen, R., Kelly, K.A., Souetre, E., Schultz, P.M., Starz, K.E.: Phase-shifting effects of bright morning light as treatment for delayed sleep phase syndrome. *Sleep.* **13**, 354–361 (1990)
23. Figueiro, M.G., Rea, M.S.: Lack of short-wavelength light during the school day delays dim light melatonin onset (DLMO) in middle school students. *Neuro Endocrinol. Lett.* **31**, 92–96 (2010)
24. Figueiro, M.G.: A proposed 24 h lighting scheme for older adults. *Light. Res. Technol.* **40**, 153–160 (2008)
25. Van Someren, E.J., Kessler, A., Mirmiran, M., Swaab, D.F.: Indirect bright light improves circadian rest-activity rhythm disturbances in demented patients. *Biol. Psychiatry.* **41**, 955–963 (1997)
26. Riemersma-van der Lek, R.F., Swaab, D.F., Twisk, J., Hol, E.M., Hoogendijk, W.J.G., Van Someren, E.J.W.: Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities: a randomized controlled trial. *J. Am. Med. Assoc.* **299**, 2642–2655 (2008)
27. Viola, A.U., James, L.M., Schlenger, L.J., Dijk, D.J.: Blue-enriched white light in the workplace improves self-reported alertness, performance and sleep quality. *Scand. J. Work Environ. Health.* **34**, 297–306 (2008)
28. Figueiro, M.G., et al.: Circadian-effective light and its impact on alertness in office workers. *Light. Res. Technol.* **51**, 171–183 (2019)
29. Van Bommel, W.J.M.: Non-visual biological effect of lighting and the practical meaning for lighting for work. *Appl. Ergon.* **37**, 461–466 (2006)
30. Kronauer, R., Forger, D., Jewett, M.: Quantifying human circadian pacemaker response to brief, extended and repeated light stimuli over the photopic range. *J. Biol. Rhythm.* **14**, 500–515 (1999)

O

Online Experiments

- [Crowdsourcing of Multilingual Color Names](#)

Op Art

- [Optical Art](#)

Opacity – Extinction

- [Apparent Magnitude, Astronomy](#)

Opalescence

- [Iridescence \(Goniochromism\)](#)

Opsin Genes

- [Photoreceptors, Color Vision](#)

Optic Chiasm, Chiasmal Syndrome

Jonathan Aboshiha
The UCL Institute of Ophthalmology and
Moorfields Eye Hospital, London, UK

Synonyms

[Chiasma opticum](#)

Definition

The **optic chiasm** (from the Greek for “cross-piece”) is the anatomical location where the two optic nerves, having left the eyes, join together before then separating once more as the two distinct optic tracts, thus forming an anatomical “X.” **Chiasmal syndrome** is the name given to the group of symptoms and signs that occur together as a result of lesions affecting the optic chiasm.

Overview

Optic Chiasm

This optic chiasm measures 15 mm in width and 3.5 mm in height [1] and has several important structures located adjacent to it; below the optic

chiasm, in a bony excavation called the sella turcica, lies the pituitary gland, and either side of it are found the venous cavernous sinuses, which transmit the internal carotid artery and several important cranial nerves. The frontal lobe of the brain sits above the chiasm, and the floor of the third ventricle lies behind. The optic chiasm is significant because it is here that the visual signals coming from each retina decussate, and so from this point onward, there is representation from both eyes on both sides (i.e., left and right sides) of the visual pathway. The nasal optic nerve fibers, carrying signals from the nasal retinae (and hence temporal hemi-fields of vision), cross or decussate to the other side (contralateral) at the optic chiasm, while the temporal fibers, which carry signals from the temporal retinae (and hence nasal hemi-fields of vision), continue on the same side (ipsilateral). In anatomically normal individuals, approximately 50% of the axons (i.e., those from the nasal retina) decussate across the midline to join the uncrossed fibers traveling toward the brain from the temporal retina. These fibers exit the chiasm to form the left and right optic tracts, which carry signals from the right and left hemi-fields of vision, respectively. In certain conditions, the proportion of crossing fibers at the chiasm is significantly different from the 50% value observed in normals; in humans with albinism, a greater number of fibers from temporal retina decussate and project contralaterally [2].

Chiasmal Syndrome

Pathology in any of the key anatomical structures that lie adjacent to the optic chiasm (see above) has characteristic effects on the optic chiasm, giving rise to a unique constellation of visual signs and symptoms, collectively known as the chiasmal syndrome. Depending on the particular pathology involving these adjacent structures, and the consequent relative location of damage to the chiasm, the exact features seen in any chiasmal syndrome can help anatomically locate the site of the lesion.

A visual field defect may be the earliest sign of a lesion at the optic chiasm [3], and the classical visual field defect seen in chiasmal syndrome is the loss of the temporal hemi-fields of vision in both eyes (so-called bitemporal hemianopia). This is caused by a lesion in the body of the chiasm which damages

the crossing nasal retinal fibers. Any lesion of the chiasm (such as a tumor, vascular lesion, traumatic lesion) may cause this, and there are subtle variations in the extent of the bitemporal hemianopia that can help give clues as to the causative lesion. The exact field defect will vary according to the type of lesion and the precise individual anatomical relationship of the chiasm to the structures that surround it. The most common cause of such a field defect is a tumor of the pituitary gland [3]. The visual acuity may also be reduced, as may the color vision on testing with pseudoisochromatic plates, with a tendency to have more severe red-green deficits and milder blue-yellow losses [4].

Lesions of the anterior chiasm can cause a so-called junctional scotoma, presenting with a central field loss in one eye and a superotemporal field loss in the other – this is classically said to be due to a small bundle of inferonasal nerve fibers that briefly loop forward in the contralateral optic nerve, called Wilbrand's knee. More recently, the existence of Wilbrand's knee has been questioned as an artifactual error, relating to the fact that the cadaveric specimens on which the observation of this anatomical feature had been based had always suffered the loss of one eye before death [5], and investigations have found no clinical perimetric evidence to support the existence of this anatomical feature in human beings [6].

Lesions of the posterior chiasm may cause an uneven field loss in both eyes of the visual field on the same side, due to damage to the optic tracts (i.e., an incongruous homonymous hemianopia).

Compression of the optic chiasm may also cause atrophy of the nasal nerve fibers, with the superior and inferior nerve fibers being relatively spared – so-called “bow-tie” optic atrophy – as well as an unusual type of nystagmus (an involuntary movement of the eyes) called “seesaw” nystagmus [7]. Loss of the temporal visual fields can lead to postfixation blindness and the inability to keep the eyes both facing in the same direction, a phenomenon known as hemi-field slide [8].

The optic chiasm also lies close to a venous space (the cavernous sinus) where the nerves that control eye movement pass – lesions in this area can damage these nerves and cause further problems with the eyes moving together, resulting in diplopia (double vision).

Anyone presenting with the above features of chiasmal syndrome should be investigated with urgent neurological imaging, in order to promptly diagnose and, if necessary, treat a potentially life-threatening cause.

Cross-References

- [Pseudoisochromatic Plates](#)
- [Visual Appearance](#)

References

1. Parravano, J.G., Toledo, A., Kucharczyk, W.: Dimensions of the optic nerves, chiasm, and tracts: MR quantitative comparison between patients with optic atrophy and normals. *J. Comput. Assist. Tomogr.* **17**(5), 688–690 (1993)
2. Hoffmann, M.B., et al.: Organization of the visual cortex in human albinism. *J. Neurosci.* **23**(26), 8921–8930 (2003)
3. Foroozan, R.: Chiasmal syndromes. *Curr. Opin. Ophthalmol.* **14**(6), 325–331 (2003)
4. Yates, J., Diamantopoulos, I., Daumann, F.: Acquired (transient and permanent) color vision disorders. In: Menu, J. (ed.) *Operational Color Vision in the Modern Aviation Environment*. North Atlantic Treaty Organization Research and Technology Organization A, Neuilly-Sur-Seine Cedex (2001), 1
5. Horton, J.C.: Wilbrand's knee of the primate optic chiasm is an artefact of monocular enucleation. *Trans. Am. Ophthalmol. Soc.* **95**, 579 (1997)
6. Lee, J.H., et al.: Wilbrand's knee: does it exist? *Surg. Neurol.* **66**(1), 11–17 (2006)
7. Unsöld, R., Ostertag, C.B.: Nystagmus in suprasellar tumors: recent advances in diagnosis and therapy. *Strabismus.* **10**(2), 173–177 (2002)
8. Kirkham, T.: The ocular symptomatology of pituitary tumours. *Proc. R. Soc. Med.* **65**(6), 517 (1972)

Optical Art

Zena O'Connor
Design Research Associates, Sydney, NSW,
Australia

Synonyms

[Op art](#); [Retinal art](#)

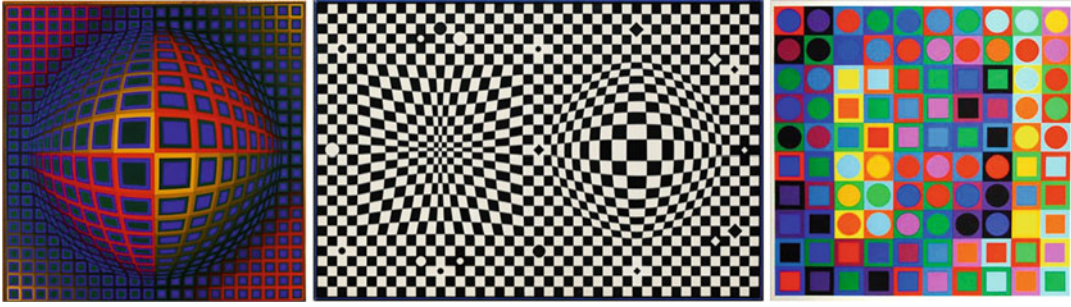
Definition

Optical art (or op art) was an art movement that emerged in the mid-twentieth century as an extension of the pop and conceptual art movements. It exploits the mechanics of human visual perception to create imagery in the viewer's mind. Op artists used line, shape, flat planes of color, and a two-dimensional format to create works that appear three dimensional and occasionally convey a sense of movement.

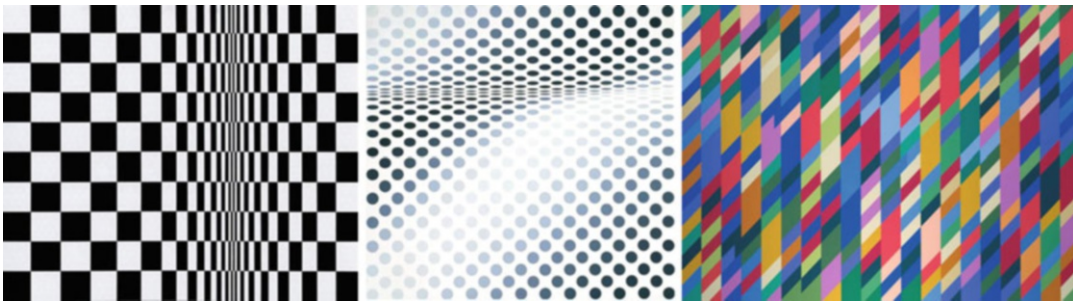
Overview

Optical or op art emerged as an extension of the pop art and conceptual art movements, both of which represented the transition from modernism to postmodernism, during the middle of the twentieth century. From a theoretical perspective, modernism rejected the accepted and established art forms but eventually became formulaic, prescriptive, and predictable, while in response, postmodernism became characterized by a more pluralistic, eclectic, diverse, and somewhat more unpredictable approach to art.

Op art “exploits the workings of perception to create a virtual reality in the viewer's mind” ([1], p. 178). Op artists aimed to create optical illusions using line, flat planes of color, and a two-dimensional format to create works that appear three dimensional and occasionally convey a sense of movement. Characterized by an emphasis on geometry, simple forms, and a strategic color ► [palette](#) (often black and white or strong contrasting colors), op art also represented “geometric precision, emblematic not only of the Space Age but... (also the) technological revolution” of the 1960s ([2], p. 73). Color and light–dark contrast are dominant visual elements in op art and used strategically to create three-dimensional visual illusions from two-dimensional formats. In addition, the strategic use of color was applied to create a sense of movement across the canvas or painted surface, and in this way, op art paintings represent outcomes that are more than the sum of their parts in terms of applied paint on surface.



Optical Art, Fig. 1 Works by Victor Vasarely



Optical Art, Fig. 2 Works by Bridget Riley

Op artists often embedded known illusions or optical effects into their works such as simultaneous contrast and colored after-images, the Hering illusion, café wall illusion, scintillating dot illusion, and the Muncker illusion.

Although she does not like the title “op art,” Bridget Riley’s works are closely associated with this movement as well as “the spirit of the mid-sixties”; and Victor Vasarely’s works are reflect the visual aesthetics of the op art movement ([2], p. 72).

Victor Vasarely (1906–1997)

Op art works by Hungarian-French artist Victor Vasarely feature the illusion of three dimensionality, and this is achieved by the strategic placement of geometric shapes, color, and contrast upon the canvas. Vasarely first started to experiment with optical patterns in the 1930s and his works were included in the Museum of Modern Art’s exhibition: *The Responsive Eye*. This exhibition, which also included works by Bridget Riley, featured

works that which were considered “less as objects to be examined than as generators of responses in the eye and mind of the viewer” ([3], p. 1). The works included in this exhibition featured works that used optical illusions, simultaneous contrast, and afterimages to produce “new kinds of subjective experiences” ([3], p. 1). As a prolific painter, key works by Vasarely include *Vega Nor* (1969), *Vega III* (1957), and *Marc Positive* (1968) (Fig. 1).

Bridget Riley (Born 1931)

Works by English painter Bridget Riley feature the illusion of three dimensionality, and this is achieved by the careful placement of variations in shape as well as color or light–dark contrast. In addition, Riley’s works often feature a sense of movement, and this is achieved by the strategic placement of color around the canvas, causing the viewer’s eye to be drawn from dominant colors and contrasts around the canvas. Figure 2 features Riley’s *Movement in Squares* (1961), *Hesitate* (1964, Tate), and *Nataraja* (2000, Tate).

The art movements of op art as well as minimalism and pop art had a strong influence on popular culture as well as fashion design in the late 1950s and 1960s [4–8]. Design output by Mary Quant, André Courreges, Pierre Cardin, and Terence Conran and firms such as Marimekko are imbued with design characteristics drawn from op art and pop art. In addition, op art influenced additional diverse sectors within applied design including graphic design and an early example is the International Wool Secretariat logo by Francesco Saroglia (1965).

Informative videos about op art include:

- Andrew Graham-Dixon on Bridget Riley video (6:30 min), <http://www.youtube.com/watch?v=Z1lQCTunGxg>
- Victor Vasarely's optical art (3:38 min), https://www.youtube.com/watch?v=UrfYC_ieI3k
- Op Art Avon TV advertisement, 1960s (1:58 min), <https://www.youtube.com/watch?v=F4WuBAOFaDk>

An artist who often harnessed illusions or optical effects into his works was Venezuelan Carlos Cruz-Diez (1923–2019). A number of small and large scale works by Cruz-Diez feature simultaneous contrast or colored after-images, and two examples include *Induction du Jaune, Série Nov 06/5* (2006) and *Induction Chromatique à double fréquence* (1993). Simultaneous contrast occurs when the perception of a color is influenced by surrounding color and colored after-images occur when color arrangement and viewing time prompts an illusion of color.

More recently, artists like Olafur Eliasson (b. 1967) have created installations and artworks that feature similar optical art effects often on a larger scale. For example, Eliasson's "Room for one color" (1997) features an immersive environment saturated with yellow light. The effect prompts a range of perceptual effects for viewers, notably changing the perceived color of objects and creating a strong after-image color on leaving the art installation [9].

Cross-References

- [Color Contrast](#)
- [Palette](#)
- [Texture Measurement, Modeling, and Computer Graphics](#)
- [Visual Illusions](#)

References

1. Bird, M.: *One Hundred Ideas that Changed Art*. Laurence King, London (2012)
2. Sandbrook, D.: *White Heat: A History of Britain in the Swinging Sixties*. Little, Brown, London (2006)
3. MoMA (Museum of Modern Art, New York): MoMA press release – the responsive eye. http://www.moma.org/pdfs/docs/press_archives/3439/releases/MOMA_1965_0015_14.pdf?2010 (1965). Accessed 10 Mar 2010
4. Garner, P.: *Sixties Design*. Taschen, Köln (1996)
5. Glancey, J.: *Modern: a Portfolio of Contemporary Interior Design Styles*. Mitchell Beazley, London (2000)
6. McDermott, C.: *Twentieth-Century Design*. Viking, London (1997)
7. Raizman, D.: *History of Modern Design*. Laurence King, London (2003)
8. Sparke, P.: *A Century of Design: Design Pioneers of the 20th Century*. Mitchell Beazley, London (1998)
9. Eliasson, O.: Room for one color. <https://olafureliasson.net/archive/artwork/WEK101676/room-for-one-color>. Accessed 21 Sept 2020

Optoelectronic Light Source

- [Light-Emitting Diode, LED](#)
- [Light-Emitting Diode, OLED](#)

Origin of Boundaries

- [Mechanisms of Color Categorization](#)

Origin of Prototypes

- [Mechanisms of Color Categorization](#)

Ostwald, Friedrich Wilhelm

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Wilhelm Ostwald was born on September 2, 1853, in Riga, the capital city of Latvia, to parents of German descent. He studied chemistry at the University of Dorpat (now Tartu, Estonia) where he received his PhD in 1878 and lectured at the Polytechnicum in Riga. In 1887, he moved to the University of Dresden in Germany where he remained until his early retirement in 1906. He was active in many fields, including philosophy, is considered to be one of the founders of physical chemistry, and received the Nobel Prize for the discovery of chemical catalysis in 1909. Ostwald is the author of 45 textbooks and over 1000 articles. In his spare time, he was a talented painter. In 1905/1906, he spent several months in the Boston area giving courses in philosophy, physical chemistry, as well as techniques of painting at Harvard University, MIT, and the Lowell Institute, where he met Munsell and learned of the latter's early development of the Munsell color system. After

he left the University of Dresden, he moved to his nearby countryseat in Grossbothen where he spent the rest of his life primarily working on color theory and developed a large color order system and a theory of color harmony. Ostwald died on April 4, 1932 Grossbothen [1].

Major Accomplishments/Contributions

In the total field of science, Ostwald saw the place of the science of color to be in psychological science [2]. He was not only fully aware of the work of Maxwell, Grassmann, and Helmholtz but was convinced that also Hering had made important contributions toward understanding of color phenomena. His immediate predecessors had focused primarily on the relationship between lights and color experience. Ostwald made significant contributions to the understanding of object colors. His main contributions in the color field are briefly mentioned below:

Color order system: Ostwald developed a color order system that combined psychological and psychophysical knowledge, resulting in *Grosser Farbenatlas* of 2500 systematically ordered samples published in 1917/1918 (Fig. 1). A reduced version of 600 samples was available in 1920. An American version of the system was published in the 1940s as *Color Harmony Manual* [3].



Ostwald, Friedrich Wilhelm, Fig. 1 View of the double cone model of Ostwald's color atlas

Distinction between unrelated and related colors: Helmholtz and others of his time investigated color primarily in terms of spectral light. Ostwald demonstrated and clarified the difference between unrelated and related colors, the latter experienced from viewing objects in varying surroundings. He showed that colors such as gray, brown, olive, and others exist in related form only [2].

Nonlinear relationship between wavelength differences and perceived hue differences: Ostwald demonstrated the highly nonlinear relationship when establishing the hue circle of his color atlas, showing that there are two spectral regions where perceived hues change rapidly as a function of changes in wavelength, while near the beginning, middle, and end of the spectrum, the frequency of change is much reduced [4].

Farbenhalb/Vollfarben: Ostwald demonstrated that idealized complementary object colors have reflectance in half of the spectrum in one case and the other half in the other case, with the transition wavelength varying as a function of hue. He used the term *Farbenhalb* (half of the spectrum) for this situation. He used a graphic format with wavelength on the horizontal and reflectance between 0 and 1 on the vertical axis. Object colors represented by *Farbenhalb* he named *Vollfarben* (full colors), object colors of the highest saturation possible for a given hue, occupying spectral ranges with either one or two transitions in the

spectrum. Full colors are the optimal object colors of a given hue at the lightness level at which they have highest saturation, varying by hue. In 1920, Schrödinger offered a mathematical proof of the concept of optimal object colors that included Ostwald's *Vollfarben* [4].

Metamerism: In 1918, Ostwald demonstrated the existence of metameric object colors and introduced the term *metamerism* [4].

Textbooks on color: During his lifetime, Ostwald published several volumes on color. The most important ones are part of a planned five-volume work *Die Farbenlehre* (The Science of Color), of which only two volumes were published during his lifetime: Vol. I *Mathetische Farbenlehre*, 1918 (ordering of color), and Vol. II *Physikalische Farbenlehre*, 1919 (color physics).

References

1. Ostwald, W.. www.wilhelm-ostwald.de. Accessed 3 Mar 2015; 3 June 2015
2. Ostwald, W.: Beiträge zur Farbenlehre. Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich Sächsischen Gesellschaft der Wissenschaften **23**(3), 365–572 (1917)
3. Jacobson, E., Ostwald, W.: The Color Harmony Manual and How to Use It. Container Corp. of America, Chicago (1942)
4. Ostwald, W.: Die Farbenlehre. Physikalische Farben, vol. 2. Unesma, Leipzig (1918)

P

Paint

- [Pigment, Ceramic](#)

Painted Dress

- [Art and Fashion Color Design](#)

Painter's Color Model

- [RYB Color](#)

Paints

- [Inorganic, Hybrid, and Functional Pigments](#)

Pale Colors

- [Pastel Colors](#)

Palette

Roy Osborne
Color Research and Application, The Colour
Group of Great Britain, London, UK

Synonyms

[Color board](#); [Color range](#); [Color selection](#)

Definition

1. Board on which paints are set in order to mix them before applying to a painting
2. The color range or selection used in a composition, art work, or design object

Introduction

In its narrowest sense, a painter's palette is a rigid board on which a set of paints is laid out and adjacent to which is an area where its ingredients can be mixed before application to a canvas, board, or other support. In its widest sense a palette represents any range of colors an artist or designer has chosen for use in any specific project. For some tasks this may simply mean black and white with various greys mixed from them; for others it can

mean an array of dozens of colors, with others mixed from them. For centuries, the palette has been an essential item in every painter's studio. Since the appearance of the Quantel Digital Paintbox (1981), many artists have replaced the paintbrush with the computer mouse, though the term "palette" is still retained to refer to the range of colors available for introduction into the visual image.

Palette, as a Board for Arranging and Mixing Paints

European palettes were traditionally made of hard wood (such as mahogany or walnut) and weighted or shaped in order to balance comfortably on the wrist or forearm of the painter while working. From about 1400, a number of portraits show small wooden palettes with handles either laying flat on a table or propped up beside a set of brushes, commonly one for each color. The latter has corners to stop them rolling over and also imply viscosity in the paint being used. For larger works, and less viscous paint, shallow bowls of single colors and large brushes were used. After about 1500, for easel paint in either tempera or oil, a palette large enough to accommodate about a dozen colors is pierced in one corner by a thumb hole. After about 1600, large curved palettes that rest on the forearm become general and define a design that remained standard into the twentieth century, though smaller, rectangular palettes that fitted into the lids of paintboxes also continued to be used. Throughout the ages, many other materials have been employed to hold and mix paints, including shells, porcelain, glass, metal, plastic, or bone, as in the case of the scapula palettes of prehistoric times. Metal trays or containers are best for wax or encaustic painting (requiring heat), and glass jars are useful for fresco painting. Shallow, cupped palettes are suited to dilute egg tempera and watercolor paint, whereas flat boards are ideal for holding small quantities of oil, alkyd, and polymer paints.

Palette as a Color Range or Selection

The common function of a palette (whatever its size and format) is to reveal the selection of

chosen colors at a glance and assist the painter in working with a degree of speed and efficiency. According to Giorgio Vasari (1568), Lorenzo di Credi (a fellow-pupil of Leonardo in the studio of Andrea del Verrocchio) "made on his palettes a great number of colour mixtures, so that they went gradually from the lightest tint to the darkest, with exaggerated and truly excessive regularity, so that sometimes he had 25 or 30 on his palette, and for each of them he kept a separate brush." More often, color mixing was kept to a minimum, and colors employed in the glorification of gods were kept as pure and unsullied as possible.

Historical Development of Palette

Of the many reference sources that document the historical variation and content of artists' palettes, among the more informative are Letalle's *Les palettes d'artistes* [1], Speed's *The Science and Practice of Oil Painting* [2], Schmid's *The Practice of Painting* [3], Bazzi's *Abecedario pittorico* [4], and Birren's *History of Color in Painting* [5].

Ancient Times

The limited range of the pigments used by the earliest painters can be ascertained from cave paintings from about 15,000 BC that have survived in northern Spain and western France. The palette employed is known to have included red ochre (hematite), yellow ochre (limonite), stibnite (grey antimony sulphide), and kaolin (white aluminium silicate). Remarkably, this four-pigment palette, known to the Greeks as the *tetrachromatikón*, not only survives in Europe but is found widely scattered throughout the globe, being a common feature in the traditional painting by the Polynesian artists of the Pacific and New Guinea, by many African tribes, and by the Australian Aborigines. Reds, greens, and yellows from vegetation may have been used but generally rapidly fade.

Ample evidence remains of the contents of the ancient Egyptian palette. In addition to fugitive dyestuffs, the Egyptians employed gypsum (calcium sulfate), white lead (lead carbonate), red lead (lead oxide), realgar (arsenic disulfide), orpiment (arsenic trisulfide), malachite green and

azurite blue (two forms of copper carbonate), Egyptian blue (crystalline copper silicate), and various carbon blacks, including lamp black and bone black. Egyptian Blue has been identified in painted decorations from about 2500 BCE onward, and its manufacture combined chalk or limestone with sand and a copper mineral, such as azurite, fired in a kiln at precise temperatures. Lapis lazuli was used in jewellery but is unlikely to have been employed as a pigment at this time.

In ancient Greece, the practice of intermixing the four earth colors was credited to the painters Polygnotus and Anaglaophon. The reliance on a palette of red ochre, yellow ochre, chalk white, and black (commonly from burnt wine lees) caused Pliny to remark, in his *Natural History* (37.12): *Quattuor coloribus solis immortalia illa opera fecere* (“Four colours alone to make their work immortal”). Apelles was acknowledged as the principal advocate of the tetrachromatikón, from which all human complexion colors could be obtained, plus the browns of hides and fur, and even the dull greens of foliage, when yellow ochre was mixed with bluish black. Additionally, Pliny writes of Apelles applying varnish to his paintings that “caused radiance in the brightness of all the colours and protected the painting from dust and dirt.” Pliny also makes a distinction between the “austere” colors (*austeri*) of the “four-earth” palette and vivid or “flowery” colors (*floridi*). Among the latter are red lead (minium), purpura lake (purpurissum), dragonsblood (cinnabaris), Armenian blue (azurite), chrysocola, and Indigo.

Greeks and Romans used much the same palette, adapted from the Egyptians, and the only significant pigment introduced by the Romans was cinnabar (natural mercuric sulfide).

Medieval Age

After the fall of the Roman empire (about 400 CE), few surviving records refer to the constitution of artists’ palettes, owing perhaps to illiteracy or a desire to keep certain practices secret, or simply because processes were so well known that no one felt the need to write them down. Commissions and contracts might specify which pigments were to be used for particular artworks,

usually with the aim of preventing the substitution of inferior colorants for precious ones similar in color, such as red lead for vermilion, or azurite for lapis lazuli, which at its finest was worth its weight in gold. The English word “guarantee” comes from the Italian term *garanza*, referring to genuine madder dyestuff. Where instructions for selecting raw materials, preparing paints, and the correct methods of using them were recorded, it was usually in the form of secret recipe books known as formularies. One of the most informative of these is *De diversis artibus* (“On diverse arts”) by the monk Theophilus (Roger of Helmarshausen), who also describes a formula for preparing artificial vermilion and whose *Schedula diversarum artium* also describes procedures for illuminating manuscripts.

An unusually complete formulary that has survived from the early Renaissance was compiled about 1400 by the Tuscan artist Cennini Cennini. Whereas Leon Battista Alberti’s treatise on painting (1435) is largely theoretical, and refers only briefly to color mixing, Cennini’s contains over 150 short chapters covering numerous aspects of workshop practice. It describes, for example, a relatively new yellow made from lead and tin oxides heated to 800°. Cennini calls it *giallorino* and describes it as heavy and permanent and useful for painting foliage and grass. In English, it became known as massicot (and subsequently lead-tin yellow) and was sometimes glazed with Indigo for depicting foliage.

Medieval artists also learnt that palettes could be extended when raw earths were roasted or calcined. Hence, yellow ochre was transformed into a reddish pigment variously called burnt ochre, light red, or English red; raw sienna and raw umber became burnt sienna and burnt umber, the latter useful for flesh tints. During the Middle Ages, a typical workshop palette might therefore include white lead, kaolin, gypsum, massicot, orpiment, yellow ochre, burnt ochre, red lead, raw sienna, burnt sienna, raw umber, burnt umber, red ochre and red bole, cinnabar or vermilion, madder lake, brazilwood lake, green earth, sap green, verdigris, malachite green, azurite blue, flower of woad, genuine ultramarine, lamp black, bone black, and vine black.

Renaissance and Beyond

Though Hubert and Jan van Eyck were not the inventors of oil painting, they are acknowledged (e.g., by Filarete) as perfecting its use in pictorial painting, together with Rogier van der Weyden. Antonello da Messina is often credited with introducing the medium into Italy, and by the end of the 1400s, oil painting was commonplace throughout Europe. Magnificent frescoes continued to be painted throughout the 1500s, notably by Raphael, Michaelangelo, and Annibale Carracci, but working from a handheld palette distributed with oil colors became almost universal practice from the mid-sixteenth to the mid-twentieth century, when North American abstract artists in particular began making large-scale abstract works. At the same time, commercial paints, alkyds, and acrylic paints stimulated new and bolder methods of working. Notable exceptions were the various schools of watercolor painters, who devised small, portable, and often ingenious palettes for outdoor sketching and made of either wood or lightweight metal.

Baroque and Classicism

With the decline in large commissions from the church, courts, or guilds, and the rise of the merchant class, it became more difficult for aspiring artists to become pupils and apprentices in large-scale workshops. To compensate for this, from the late 1600s onward, a series of books began to appear that offered tuition in the basic skills of assembling and applying practical palettes. Among the first was William Salmon's *Polygraphice* (1672), followed by Claude Boutet's *Traité de la mignature* (1673), modified through dozens of editions. Alternatively, a student might learn from a single master but had better chance of a more rounded education if he could enroll at one of the many new academies, though even in the 1700s, these were few and far between. Italian painting continued to be much admired, and writers often bemoaned the fact that artists like Titian had not passed their "secrets" of color-mixing down to later generations, though Lomazzo's *Trattato dell'arte de la pittura*

(1584) had preserved some of the late-Renaissance practices.

The glazing of three primary colors, demonstrated in color circles, was included in Moses Harris' *Natural System of Colours* (1766), offering a new concept of tricolor mixing, and extended in Johann Lambert's *Farbenpyramide* (1772) and James Sowerby's *New Elucidation of Colours* (1809) and culminating in Philipp Otto Runge's three-dimensional *Farben-Kugel* (1810). As the science of chemistry accelerated throughout the early nineteenth century, many new artificial pigments were added to the artists' palette, notably cobalt blue (1804), chrome yellow (1818), cadmium red and yellow (1817), Guimet's artificial ultramarine (1826), and chromium oxide green (1838).

Nineteenth Century

In excess of 50 instructional books on how to use color palettes were published during the 1800s, though few influenced the development of fine art painting as such. One exception is Ogden Rood's *Modern Chromatics* (1879), responsible (in its French edition of 1881) for clarifying Chevreul's earlier explanations of optical color mixing. By the mid-1880s, numerous artists (including Georges Seurat, Vincent van Gogh, and Camille Pissarro) were laying out their colors in spectral sequence, with white available for tinting.

Twentieth Century

Over the next 30 years, almost every progressive painter experimented with some form of pointillism, and its novel way of applying unmixed color directly from the palette to the canvas (notably in early twentieth-century French and German expressionism) led to the liberation of color as a pictorial element in its own right, and one no longer restricted to the traditional task of describing forms in the natural or built environment.

During the early years of the twentieth century, avant-garde artists became less concerned with the subtleties of matching their colors to external

objects as with the relationships between the colors arrayed on the palette. Such a preoccupation was reinforced first by the publication of Albert Munsell's *A Color Notation* (1905) and then of Wilhelm Ostwald's *Die Farbenfibel* (1916), which was highly publicized throughout European schools of art and design. One limitation of such systems, and others like them, was their emphasis on opaque colors, whereas many artists (including Paul Klee) typically experimented with translucent oil or watercolor glazes. Additionally, as if in reaction to the color explosion of the Impressionist generation, the Cubists reverted to the ancient tetrachrome palette, with Georges Braque, Pablo Picasso, and even Marcel Duchamp, restricting their palettes to white, black, red and yellow ochre, and subdued green earth. No theorists or writers of art manuals influenced their decision, yet a general reversion to classicism held sway until the mid-1950s, through various realist revivals, until vivid palettes came to the fore again in the 1960s, with the extravert and confident art movements variously labeled Pop Art, Op Art, and ► [Color Field Painting](#).

Cross-References

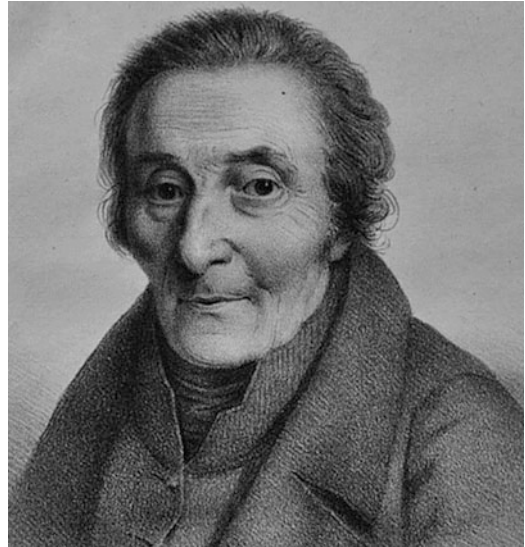
- [Ancient Color Terminology](#)
- [Chevreul, Michel-Eugène](#)
- [Color Field Painting](#)
- [Colorant, Natural](#)
- [Impressionism](#)
- [Inorganic, Hybrid, and Functional Pigments](#)
- [Op Art](#)

References

1. Letalle, A.: *Les Palettes d'Artistes*. E. Sansot, Paris (1912)
2. Speed, H.: *The Science and Practice of Oil Painting*. Chapman and Hall, London (1924)
3. Schmid, F.: *The Practice of Painting*. Faber & Faber, London (1948)
4. Bazzi, M.: *Abecedario Pittorico*. Longanesi, Milan (1956). Translated as *The Artists' Methods and Materials*. John Murray, London (1960)
5. Birren, F.: *History of Color in Painting*. Van Nostrand Reinhold, New York (1965)

Palmer, George

Rolf G. Kuehni
Charlotte, NC, USA



Louis Fehr: George Palmer, ca. 1820

Biography

George Palmer, also known as George Giros de Gentilly named Palmer, was an English dye chemist, color theorist, inventor, and soldier. According to his obituary, Palmer was born ca. 1746 on a ship, to English Catholic parents. Due to the eighteenth-century restrictions on activities of English Catholics, Palmer lived a double life between England and France. Nothing is known about his early years. Circa 1775, he introduced a solution of tin as a new mordant for the dyeing of wool fabrics in Louviers, France, using the last name Giros de Gentilly [1, 2]. In 1777, located in London, he published the book *Theory of Colours and Vision*, a French edition of which was published in the same year in Paris, translated by Palmer's friend Denis-Bernard Quatremère d'Isjonval, at the time active in the textile manufacturing facility Disjonval in Sedan,

France, owned by his family [3, 4]. In the same year, Palmer also invented a fawn-colored dye in London [5]. In 1781, J. H. Voigt of Gotha, Germany, editor of *Magazin für das Neueste aus der Physik und Naturgeschichte* (Journal for the Latest Physics and Natural Sciences News), describes meeting with Giros von Gentilly and the latter's conjectures about color blindness [6]. In 1785, Palmer, living in Paris, had *Lettre sur les moyens de produire, la nuit, une lumière pareille à celle du jour* (Letter Concerning the Means of Producing at Night a Light Equal to Daylight) published [7] describing the modification of oil lamplight with a blue glass mantel, a technology that became fashionable for a time. In 1786, Palmer published *Théorie de la lumière, applicable aux arts, et principalement à la peinture* (Theory of Light Applicable to the Arts, Principally to Painting) [8]. Toward the end of that decade, likely as a result of the French Revolution, Palmer became a mercenary soldier in the Corps of Engineers, at different times for Sweden, Austria, and Russia, reaching the rank of major, as described in his obituary [9]. For a time in the early nineteenth century, he lived near Leipzig in Germany where he reported on four technical inventions, one of which being a fire-extinguishing powder, a demonstration of which was reported in a local newspaper. In 1811, Palmer moved to Copenhagen into retirement and died there destitute in 1826 [10].

Major Accomplishments/Contributions

Palmer made lasting contributions to the development of color science by being the first to propose that there are three different mechanisms in the human eye that account for color vision: "The superficies of the retina is compounded of particles [light sensors] of three different kinds, analogous to the three rays of light; and each of these particles is moved by its own ray" [4]. This statement has proved true in regard to the number of different daylight sensor types, the cones, in the human eye, if not in regard to the claim of three

kinds of light. Thirty-five years later, a similar statement was made by the eminent physicist Thomas Young [11].

Voigt, in his report on Giros von Gentilly, describes him as having stated that color blindness arises if one or two of the three kinds of "particles" in the retina are inactive, a statement found to be valid [6]. In 1786, Palmer provided a hypothesis for the complementary nature of the successive contrast effect by stating that it is due to fatiguing of one or two of the light sensor types, an explanation that continues to be accepted as valid, as does his conjecture that the different kinds of sensor take different times to recover upon exposure to strong light [8].

References

- Berthollet, C.L., Berthollet, A.B.: Elements de l'art de la teinture, 2nd edn. Firmin, Didot, Paris (1804)
- Macquer, P. J.: Mémoire des Sieurs Maille frères et de Lafosse manufacturiers de drap a Louviers, Aug. 3, 1781. Document No. 5315, Box F12/1334B, Archives Nationales de France, Paris (1781)
- Palmer, G.: Theory of Colours and Vision. Leacroft, London (1777)
- Palmer, G.: Théorie du couleur et de la vision (trans. Quatremère d'Isjonval, D.-B.). Pissot, Paris (1777)
- Palmer, G.: Letter dated June 19, 1793, Letter from Lord Hervey including two letters in French and two samples by G. Palmer about his yellow dye, Royal Society for the Encouragement of Arts, Trades, and Manufactures, London, Ref. No.: PR-MC/105/10/452 (1793)
- Voigt, J.H.: Des Herrn Giros von Gentilly Muthmassungen ueber die Gesichtsfehler bey Untersuchung der Farben. Magazin für das Neueste aus der Physik und Naturgeschichte. **1**(2. St.), 57–61 (1781)
- Palmer, G.: Lettre sur les moyens de produire, la nuit, une lumière pareille à celle du jour, Paris (1785)
- Palmer, G.: Théorie de la lumière, applicable aux arts, et principalement à la peinture. Hardouin et Gattey, Quinquet, Paris (1786)
- Thaarup, F.: Dagen newspaper, Copenhagen, No. 66 (March 18), No. 68 (March 21) and No. 70 (March 23) (1826)
- Kuehni, R.G.: Which George Palmer? Color Res. Appl. **34**, 5–9 (2009)
- Young, T.: The Bakerian Lecture, On the theory of light and colours. Philos. Trans. R. Soc. Lond. **92**, 12–48 (1802)

Parafovea

Masato Sakurai
Shizuoka Institute of Science and Technology,
Fukuroi, Japan

Synonyms

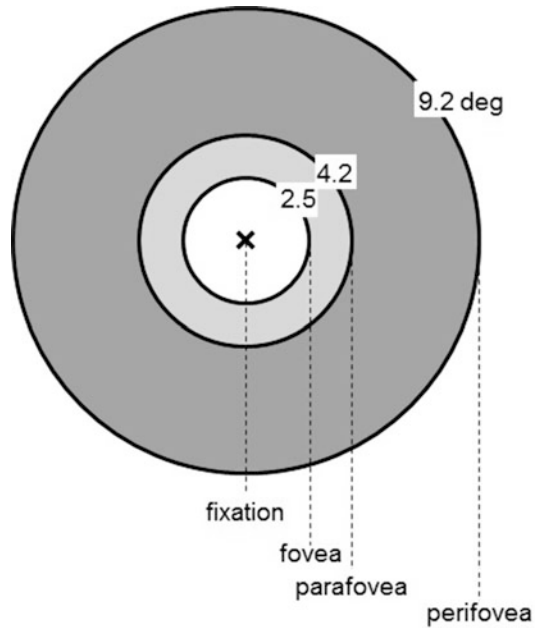
Parafoveal vision

Definition

The parafovea may refer to a position at a visual angle of approximately 4° from the normal fixation point or to the region extending to that angle from the fovea. It is compositionally distinguished from the fovea by its lower cone and ganglion cell densities and by its higher rod density. As may be expected, these differences affect its visual perception characteristics. Its spatial contrast sensitivity, visual acuity, and other visual attributes are lower than those of the fovea, but, as shown by its critical fusion frequency (CFF), its temporal frequency resolution extends to higher frequencies than that of the fovea. Its color perception is essentially the same as that of the fovea, but under low luminance, it becomes similar to tritanopic diachronic perception. These characteristics can be inferred from the relative distribution of the three cone types in the parafoveal retina. In reading, moreover, parafoveal information contributes to line of vision movement and to fine foveal information processing.

Overview

The retinal structure in the central visual field comprises three distinctive regions. As shown schematically in Fig. 1, the parafovea lies between the fovea and the perifovea. The fovea lies in the central region extending from the fixation point to a visual angle of 2.5° radial, the parafovea surrounds it and extends out to 4.2° radial, and the perifovea extends from there to 9.2° radial [1, 2].



Parafovea, Fig. 1 Central visual field on monocular viewing [1, 2]

These demarcations are not strictly defined, but phenomena occurring approximately 4° radial from the normal fixation point are generally referred to as parafoveal.

In the parafovea, the cone density declines from the high concentration of cones in the fovea, the rod density rises from the complete absence of rods in the fovea [3], and the density of ganglion cells, which form the retinal outlet of the optic nerve, declines together with the ganglion cell layer thickness from the high density and thickness of the foveal region [4, 5]. Although the parafovea and the fovea both form part of the macula, they differ significantly in macular pigment density [6, 7]. These structural and compositional differences suggest that they affect parafoveal visual perception.

Spatial contrast sensitivity, a fundamental characteristic of visual perception, is lower in the parafovea than in the fovea. The parafoveal contrast at an eccentricity of 4° is functionally characterized as a spatial frequency band-pass filter, like that of the fovea at 0° , but with lower overall sensitivity [8, 9]. In the region most sensitive to spatial frequencies of 2–8 cycles/ $^\circ$, the parafovea

sensitivity is approximately 1/2–1/3 of the foveal sensitivity [9]. Visual acuity is also significantly lower than in the fovea. In Landolt ring visual acuity measurements, subjects with a foveal visual acuity of 2.0 exhibit a sharply lower visual acuity of 0.3 in the parafoveal region at approximately 4° [10], thus showing that a larger character size is necessary for character deciphering. It has been found that the character size threshold for deciphering increases linearly throughout the foveal region and on out to an eccentricity of 30° [11]. In temporal frequency contrast sensitivity, on the other hand, the parafoveal resolution extends to a higher frequency than the fovea, as measured by CFF (critical fusion frequency) [12].

Parafoveal color perception does not depart from that of the fovea under relatively high illuminance, but under low illuminance, it resembles tritanopic dichromatic perception (small-field foveal vision) [13]. This may be inferred from psychophysical measurements that show nearly the same L- and M-cone ratios in the fovea and the parafovea [14] but extremely low S-cone distribution in the parafovea as compared with its L- and M-cone distributions [6].

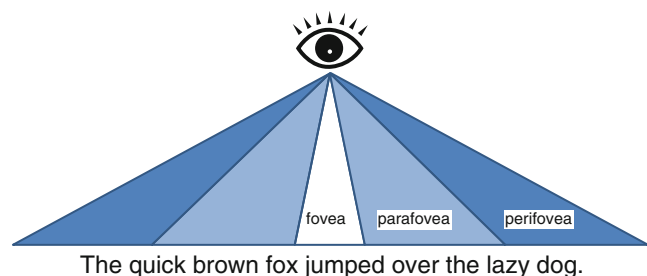
Reading experiments on parafoveal information processing have shown that efficient reading requires simultaneous parallel processing of parafoveal and foveal information [15]. Figure 2 shows the foveal, parafoveal, and perifoveal regions in reading with three characters occupying a visual angle of 1° [15]. The foveal region, which involves processing in finer detail, is narrow. It occupies only approximately seven characters, but it receives parafoveal information concerning the position of the subsequent characters, which results in a shift in the line of sight toward them by movement of the eyeball. This

has been confirmed experimentally by having subjects gaze at the fixation point and then showing them a word in the paranoveal region, which immediately elicits a response in which the subjects move the line of sight and read that word.

Three phenomena that have been observed in reading experiments concerning parafoveal information processing are known as skipping, parafoveal-on-foveal (PoF), and preview-benefit effects. The skipping effect consists of skipping every third word or so, which would presumably not be possible without identification of the word via the parafovea. The skipped words are generally short terms such as “to,” “for,” “in,” and “the,” which supports the view that a word’s lexical status contributes to the determination of whether it will be skipped.

In the PoF effect, the occurrence of a parafoveal word that is difficult to understand affects the foveal fixation and thus the field of vision. As noted above, parafoveal processing affects foveal processing, and a low-level PoF effect has been well established in experiments involving the occurrence of unfamiliar character combinations in the parafoveal region. The preview-benefit effect has been observed in experiments in which the word on the parafovea that is subsequently to be subjected to foveal processing is replaced by a related word during the saccade. If the replacement word is related to the original parafoveal word, the fixation on that word by the fovea is shorter and its processing thus faster than if the replacement word is unrelated. Recent studies have indicated that the preview-benefit effect occurs in sentences independently of the initial sentence word, mid-sentence words, and final sentence word.

Parafovea, Fig. 2 Foveal, parafoveal, and perifoveal regions in reading when three characters make up 1° of visual angle [15]



On the other hand, it is proposed that parafoveal information processing takes place in two stages: the first is an orthography-based pre-attentive stage, and the second is an attention-dependent lexical access stage by the results from a display change paradigm experiment [16]. Also, it is reported that the efficient parafoveal pre-processing in reading may require binocular vision by measurement of reading time with reduced binocular input in the parafovea [17].

These three effects and its relation for binocular vision are currently the subject of intense discussion and further findings may be expected.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Cone Fundamentals](#)
- [Tritanopia](#)

References

1. Polyak, S.L.: The retina. University of Chicago Press, Chicago (1941)
2. Osaka, N.: Psychophysical analysis of peripheral vision. Kazama Shobo Press, Tokyo (1983). (in Japanese)
3. Wandell, B.A.: Foundations of vision. Sinauer Associates, Sunderland (1995)
4. Curcio, C.A., Allen, K.A.: Topography of ganglion cells in human retina. *J. Comp. Neurol.* **300**, 5–25 (1990)
5. Iwasaki, M., Inomata, H.: Relationship between superficial capillaries and foveal structures in the human retina. *Invest. Ophthalmol. Vis. Sci.* **27**, 1698–1705 (1986)
6. Gegenfurtner, K.R., Sharpe, L.T. (eds.): Color Vision from Genes to Perception. Cambridge University Press, Cambridge, UK (1999)
7. Werner, J.S., Bieber, M.L., Scheffrin, B.E.: Senescence of foveal and parafoveal cone sensitivities and their relations to macular pigment density. *J. Opt. Soc. Am. A.* **17**(11), 1918–1932 (2000)
8. Rovamo, J., Virsu, V., Nasanen, R.: Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature.* **271**, 54–56 (1978)
9. Virsu, V., Rovamo, J., Laurinen, P.: Temporal contrast sensitivity and cortical magnification. *Vis. Res.* **22**, 1211–1217 (1982)
10. Virsu, V., Nasanen, R., Osmoviita, K.: Cortical magnification and peripheral vision. *J. Opt. Soc. Am. A.* **4**(8), 1568–1578 (1987)
11. Anstis, S.M.: A chart demonstrating variations in acuity with retinal position. *Vis. Res.* **14**, 589–592 (1974)
12. Hartmann, E., Lachenmayr, B., Brettel, H.: The peripheral critical flicker frequency. *Vis. Res.* **19**, 1019–1023 (1979)
13. Gilbert, M.: Colour perception in parafoveal vision. *Proc. Phys. Soc.* **63**, 83–89 (1950)
14. Neger, J.L., Cicerone, C.M.: The ratio of L cones to M cones in the human parafoveal retina. *Vis. Res.* **32**, 879–888 (1992)
15. Schotter, E.R., Angele, B., Rayner, K.: Parafoveal processing in reading. *Atten. Percept. Psychophys.* **74**, 5–35 (2012)
16. Angele, B., Slattery, T.J., Rayner, K.: Two stages of parafoveal processing during reading: evidence from a display change detection task. *Psychon. Bull. Rev.* **23**(4), 1241–1249 (2016)
17. Nikolov, M., Jainta, S., Blythe, H.I., Liversedge, S.P.: Binocular advantages for parafoveal processing in reading. *Vis. Res.* **145**, 56–63 (2018)

Parafoveal Vision

- [Parafovea](#)

Parallel Visual Pathways

- [Magno-, Parvo-, and Koniocellular Pathways](#)

Pastel Colors

Malvina Arrarte-Grau

Landscape and Color Designer, Environmental Color Design Study Group (AIC), Lima, Peru

Synonyms

[Pale colors](#); [Pastel crayons](#); [Pastel paintings](#); [Pastel sticks](#); [Pastel tones](#); [Soft colors](#)

Definition

The range of pastel colors, usually described as pale, is characterized by high value and low saturation. A more subjective definition of pastel colors addresses them as soft and soothing. In painting, the term pastel comes from the French *pastille* [1, p. 117], meaning small loaf or lozenge, in relation to the characteristics of the stick used as a coloring medium. The noun “pastel” also denotes the artwork produced with pastel crayons. As an adjective, “pastel tones” refer to light, pale, clean colors, with an effect similar to the one produced with pastel crayons.

The Pastel Crayon and the Pastel Technique

The pastel crayon is a compact stick used as a dry coloring medium, which combines a pure powdered pigment and a neutral binding for cohesion and plasticity. A component of calcium carbonate, such as chalk, gypsum, or pumice, is used in the preparation of the stick, resulting in various degrees of hardness and in gradations from the purest to the highly luminous. Hard pastels contain a higher proportion of binder, are sharper, and work for drawings and sketches. Soft pastels usually come wrapped in paper, for they are drier and more fragile.

The pastel technique consists of transferring small granules of color from the crayon to the support, usually textured or abrasive for good adherence. A fixative is required for protection after its application. In pastel paintings, surfaces result extremely opaque as high concentrations of pigment allow no transmission of light to the substrate, and hence, there is also no light reflected from the substrate toward the observer through the pigmented layer.

Given the directness and ease of transference of the medium, as well as its purity and similarity in color to those of the natural ► [pigments](#), pastel impressions served as the first measurement standards in colorimetry [1, p. 117].

But the effects of pastels may also be delicate and vaporous, as the ones produced in sketches and drawings. The pastel technique has the

particularity of being used in layers, with the possibility of adding lighter touches over the painted surface, to be diffused and blended mechanically for a lighter and blurred effect. In comparison to traditional pastel crayons, oil pastels have a more intense outcome.

The Pastel in Art

It is said that pastels originated in France. Leonardo da Vinci referred to them in his *Codice Atlantico* (1495) as “a new dry substance” [2, pp. 63–87]. The Venetian painter Rosalba Carriera and the French artist Quentin de La Tour represent the development of the pastel technique mainly used for portraits during the eighteenth century, with fast and realistic effects. The immediate results offered by this medium, appropriate for movement and landscape paintings, made the pastel technique the preferred art form of the ► [impressionist](#) Edgar Degas in the representation of ballet rehearsals and ballerinas. Eduard Monet used pastels for sketching landscapes scenes in rapidly changing lighting conditions, and Henry de Toulouse-Lautrec, in his color sketches of women in the Paris dance scene. The American Marie Cassat, influenced by Degas, created many of her most important works in this medium. Later, pastel tones became popular in Modern Art for their bright and luminous colors.

Pastel Colors and the Pastel Palette

Pastel colors refer to pale clean luminous nuances. The colors in a pastel ► [palette](#) are different from the colors in a pastel crayon set, for in the latter the consistency of the coloring material allows a variety that surpasses pale and luminous nuances. In subtractive mixing, pastel tones would be obtained by adding white to a base color; in this way, the colors recognized as pastel have whiteness as a common denominator. This particularity results from low chromaticness and no content of black.

In occidental culture, the pastel ► [palette](#) evokes spring, summer, and Easter and has been traditionally used in infant decoration for its

softness and gaiety. Pastel tones are also associated with the light and velvety texture of pastel paintings, as in wool and cotton, sugar icing, and powder eye makeup.

Application in Industry and Design

As in all industries, the production of materials makes the possibilities of color available to the consumer. The pastel palette has been recurrent in fashion trends in commercial wall paint. The preparation of pastel colors in paint, as opposed to that of saturated colors, allows a relatively high proportion of white binder and a lower amount of pigment. In general, it could be said that it is more economical to produce. Additionally, interiors painted in pastel colors withstand the passage of time, as they create a beneficial effect from the point of view of environmental design, granting luminosity to the space.

Pastel colors are eloquent when applied to design and frequently associate with a particular style. These typically reinforce vintage images of the world of fashion, decoration, and architecture, such as artifacts of the 1950s and Art Deco buildings, recalling places as Ocean Drive in South Miami Beach. The Miami Vice series from the 1980s brings to mind memorable lavender suits and pink trousers, from a decade when pastel tones became a trend in architecture, interior design, and product design, ranging from men's ties to motorcycles [3]. More recently, pastel Goth combines black leather accessories and dark makeup with pastel-colored wigs and tutus.

In the food manufactory, traditional sweets containing sugar and/or egg whites, such as Easter eggs, doughnuts, and French macarons are offered in appealing pastel colors. At present, around the world, in search of healthier ingredients, matcha, pitahaya, and a variety of fruits give ice cream and delicate sweets, a healthier and visually attractive appearance, resulting in amazing pastel colors.

The consistent application of pastel color in certain industries addresses its relevance in the definition of style and branding of classical objects as could be a 1955 pink Cadillac or a 1961 turquoise Volkswagen Beetle, or a more accessible mint green Vespa.

This is also the case of logos such as the Tiffany & Co. logo in the so-called Tiffany blue. As a resource in graphic design, pastel colors may be combined with the boldness of shape, taking the advantage of contrast against nuances of higher intensity. Flat design is achieved by using pastels or gray as background to selected letter types and figures, in order to create a legible, clear, and well-balanced composition. In line with that concept, the pastel palette has been used in the production of stationery, as it enhances ink hand-writing and dark lettering. It is also promoted in web design [4] through web stencils, which allow the readability of captions and pictures.

Cross-References

- Impressionism
- Palette
- Pigments

References

1. Delamare, F., Guineau, B.: *Les matériaux de la couleur*. Gallimard, Paris (1999)
2. Geddo, C.: Il “pastello” ritrovato: un nuovo ritratto di Leonardo? *Artes*. **14**, 63–87 (2008–2009). https://www.academia.edu/12454912/LEONARDO_Il_pastello_ritrovato_un_nuovo_ritratto_di_Leonardo. Accessed 31 Mar 2021
3. [https://en.wikipedia.org/wiki/Pastel_\(color\)](https://en.wikipedia.org/wiki/Pastel_(color)). Accessed 31 Mar 2021
4. <https://usabilla.com/blog/the-effects-of-pastel-colors-in-web-design/>. Accessed 31 Mar 2021

Pastel Crayons

- Pastel Colors

Pastel Paintings

- Pastel Colors

Pastel Sticks

- ▶ [Pastel Colors](#)

Pastel Tones

- ▶ [Pastel Colors](#)

Pattern Induced Flicker Colors

- ▶ [Fechner's Colors and Benham's Top](#)

Pearlescence

- ▶ [Iridescence \(Goniochromism\)](#)

Perceived Reflectance

- ▶ [Anchoring Theory of Lightness](#)

Perceived Value

- ▶ [Anchoring Theory of Lightness](#)

Perceptual Appearance

- ▶ [Appearance](#)

Perceptual Color Effects

- ▶ [Psychological Color Effects](#)

Perceptual Grouping and Color

Daniele Zavagno¹ and Olga Daneyko²

¹Department of Psychology, University of Milano-Bicocca, Milan, Italy

²Department of Psychology, Sociology and Politics, Sheffield Hallam University, Sheffield, UK

Synonyms

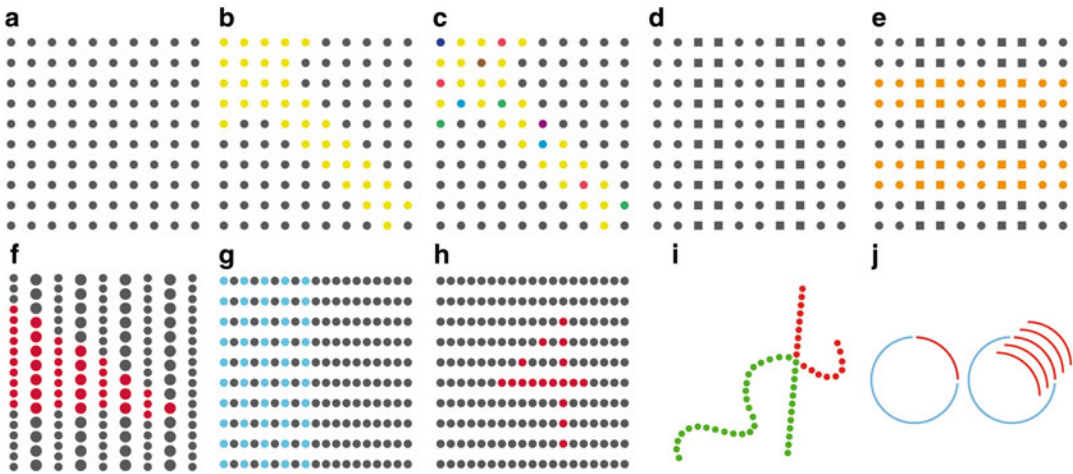
[Gestalt grouping principles \(or laws\)](#)

Definition

Perceptual grouping refers to principles by which means a set of physically discrete elements are partitioned into perceptual groups by the visual system, thus forming higher order perceptual units. Perceptual grouping was first studied by Max Wertheimer, one of the fathers of Gestalt psychology [1]. It is a core topic within the studies of perceptual organization, along with figure-ground segmentation [2]. When it comes to color, there are two aspects of perceptual grouping that need to be considered separately: the role of color in perceptual grouping; the effects of perceptual grouping on the perception of color. In the first case, the focus is on how color can direct and/or affect the perceptual organization of a visual scene. In the second case the focus is on how a color is affected by grouping processes.

The Effect of Color on Perceptual Grouping

Perceptual grouping is, along with figure-ground segmentation, a core topic in the studies of perceptual organization, i.e., of those processes that structure the sensory input into coherent units (visual objects, entities, and events), thus contributing significantly to the layout of the visual scene. From such a definition, it is clear



Perceptual Grouping and Color, Fig. 1 Grouping by color similarity. (a) Gray dots organized in a regular square lattice within which no other group or visual structure emerges. (b) By selectively introducing a change of color in some of the dots one can generate specific patterns; (c) notice that the strength of the pattern is directly influenced by the color similarity of the dots: in this lattice, 30% of the dots that form the arrow are made of colors other than yellow, and the impression of an arrow-like structure is weakened. (d) This lattice is made of round and square dots; because of shape similarity the lattice appears vertically organized in pairs of columns; (e) color similarity can overrun shape similarity, switching the perceptual organization from columns to rows. (f) This pattern would normally appear as dots organized in columns according to the factors size similarity and proximity; by introducing color, another pattern within the pattern appears. (g) Because of

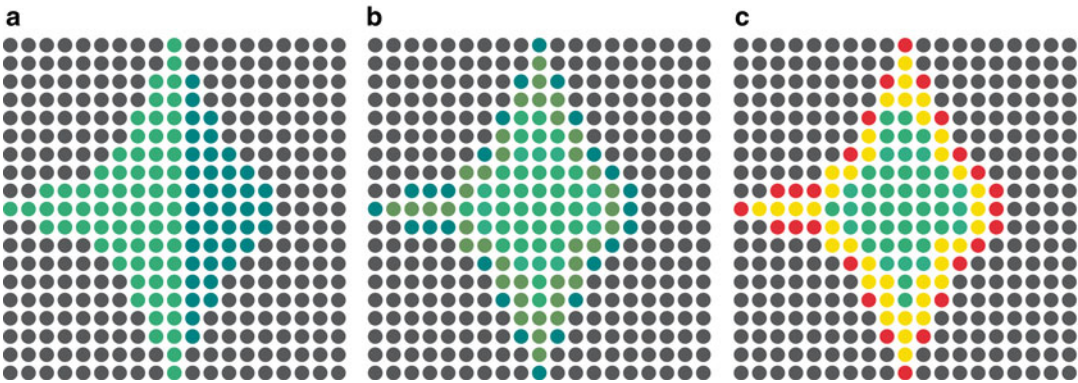
proximity, the configuration normally appears organized in horizontal rows (right side); color similarity, however, overruns the factor proximity forcing a perceptual organization in vertical columns (left side). (h) Number 4 appears in spite of proximity and good continuation; notice, however, that in this configuration and the one in (e), the red patterns can be seen as if they were under the lattice, an impression that recalls transparency. (i) Color similarity does not disrupt good continuation; instead it generates a second order group because of which the interception between the curvy line and the straight line becomes a salient feature of the configuration. (j) On the left the factor closure determines the impression of a circle with a red arch; on the right, the red arch is grouped with other red ones, masking the completion effect. Notice, however, that in this last case the light blue circle appears to complete amodally under the structure formed by the red arches

that *grouping* and *segmentation* are two faces of the same coin, as originally postulated by the fathers of Gestalt psychology Max Wertheimer, Wolfgang Köhler [3], and Kurt Koffka [4].

The main factors that regulate perceptual grouping are the following: *proximity*, *similarity*, *common fate* (similarity in motion or change in time), *good continuation*, *closure*, *prägnanz* (or *good Gestalt*, or *structural coherence*), and *past experience*. These factors are also known as Gestalt “principles” or “laws.” Color comes into play with the factor *similarity*. Figure 1a shows a regular square lattice of dots; these can be grouped in many ways to form subpatterns, yet none of the resulting patterns is specific to the lattice, nor are they durable perceptions. By selectively introducing a color difference between the dots, the visual system can be forced to organize the dots of the

lattice into stable patterns or to form larger units based on color similarity. Figure 1b, for instance, shows a yellow arrow; notice, however, how the grouping effect weakens if the dots that make the arrow in Fig. 1b show instead multiple colors as in Fig. 1c. Color similarity is a relatively strong factor that can overrun other types of similarities, such as *shape* (Fig. 1d, e) and *size* (Fig. 1f). It can also overrule a relevant factor such as *proximity* (Fig. 1g, h); however, it cannot easily overrun *good continuation* (Fig. 1i). Finally, while color differences among elements do not seem to affect closure (Fig. 1j left), grouping by color similarity can be employed to mask closure (Fig. 1j right).

As a grouping factor, color similarity is not “all or none” (Fig. 2): this makes color similarity an extremely ductile tool in art, art restoration, design, cognitive ergonomics, etc. as it can be



Perceptual Grouping and Color, Fig. 2 Grouping by color similarity is not all-or-none: in (a) the light green and the dark green dots are grouped separately in substructures that are then combined together to form the impression of a global ray fish-like structure. In (b) dots are of three different shades of green, loosely organized with brighter greens at the center and darker greens in the periphery of the ray fish-like structure. This type of grouping is commonly observed in pictorial art and can be easily appreciated in the art movement known as Pointillism. The case

represented in (c) is somewhat similar to the one shown in Fig. 1c; however, here the ray fish-like structure emerges more easily because of the limited number of colors used and the way they are ordered, with the green dots grouped in the center, the yellow ones surrounding the green dots, and the red dots delimiting the periphery. This kind of grouping can be exploited to show energy patterns (for instance, patterns of neural activation, intensity distributions, etc.)

used to enhance, impose, or create visual structure or to hinder or mask the perception of structure, form, such as in animal mimicry or in military camouflage.

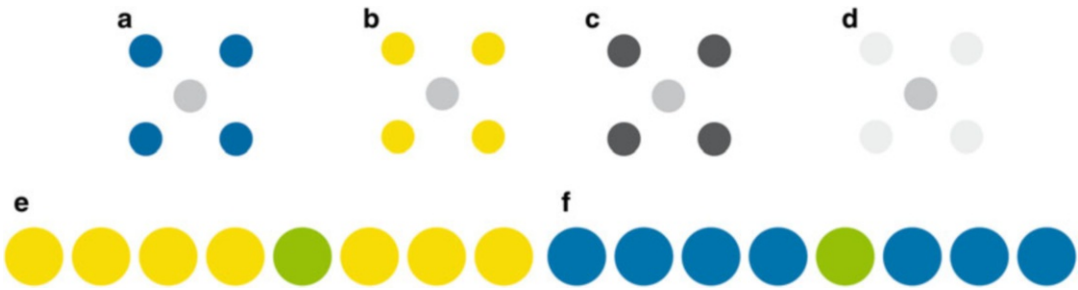
The Effects of Perceptual Grouping on Color

More problematic is the account of the effects of perceptual grouping on color. In an early paper, Fuchs [5] described situations in which perceptual grouping lead to color assimilation. Though sometimes reported by different authors (see for instance [6, 7]), older books dedicated to color perception [8, 9] curiously do not make mention to Fuchs' findings on color assimilation and grouping. The issue may depend on the fact that Fuchs' demonstrations do not always work on paper (or screen), maybe because they were mostly created by using light, shadows, projections, and reflections.

An attempt to graphically replicate one of Fuchs displays is made in Fig. 3a, b, in which two identical gray dots are placed one at the center of a square made of four blue dots and the other at

the center of a square made of four yellow dots. According to Fuchs, the gray dot surrounded by blue dots should appear bluish, while the gray dot surrounded by yellow ones should appear yellowish. The impression, however, though rather weak, seems to go in an opposite direction: that of contrast, at least as far as lightness (achromatic surface color, Fig. 3c, d) is concerned. Figure 3e, f employs proximity and good continuation to create two rows of respectively yellow and blue dots, which both incorporate an identical green dot: for those who will get an assimilation effect, the green dot grouped with the yellow dots will look somewhat more yellowish (or *warmer*) than the green dot grouped with the blue ones. It is not uncommon, however, to have observers reporting an opposite finding, or the same observer reporting switches from assimilation to contrast and vice versa.

While the perceptual outcomes emerging from the observation of Fig. 3e, f are not clear, those deriving from Fig. 3a, b recall the effects of perceptual grouping on lightness [10]: a middle gray target that is grouped with dark gray surfaces will look lighter than a middle gray target that is grouped with light gray surfaces (Fig. 3c, d).



Perceptual Grouping and Color, Fig. 3 (a, b) A graphic interpretation of one of Fuchs' displays: the central gray dot should appear bluish when surrounded by blue dots and yellowish when surrounded by yellow dots. (c, d) A typical lightness contrast effect due to perceptual

grouping and accountable in terms of the anchoring theory for lightness perception [6]. (e, f) Two straight lines of dots (proximity, good continuation), one yellow and the other blue, both including a green dot. Similar configurations can give rise to either assimilation or contrast

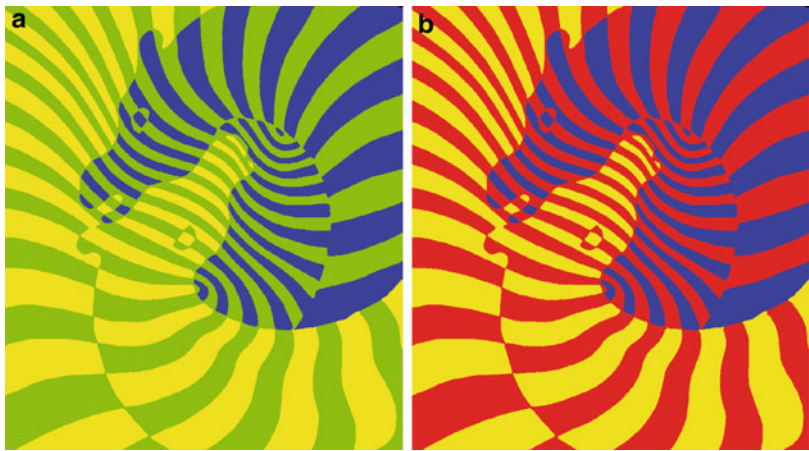
Perceptual Grouping, Belongingness, Assimilation, and Contrast

Was Fuchs completely wrong when he wrote about color assimilation as an effect of grouping elements into meaningful wholes? As mentioned earlier, his experimental setups were quite different from those commonly used nowadays, hence before any conclusion can be drawn his experimental setups need to be replicated. Moreover, Fig. 4 may be considered in support of Fuchs claims, who insists on the idea of assimilation as a reduction of differences among the parts grouped to form a Gestalten. In Fig. 4, the two configurations made of zebra-like structures, derived from a modification of a well-known graphic artwork by Victor Vasarely, show strong effects of color assimilation in which not only the "critical" color (to use Fuchs terminology – in Fig. 4a green stripes, in Fig. 4b red ones) assimilates chromatic qualities of the stripes with which they are grouped, but also the blue and yellow stripes pickup characteristics of the critical colors. For instance, in Fig. 4a the yellow stripes have a greenish tinge while in Fig. 4b they have a slight orange tinge.

Apart from the difficulties in reproducing Fuch's findings, it is clear that grouping has an effect on color perception, with a perceptual result that, however, may shift from color assimilation to color contrast. This ambiguity may depend on different factors, including a shift of weights

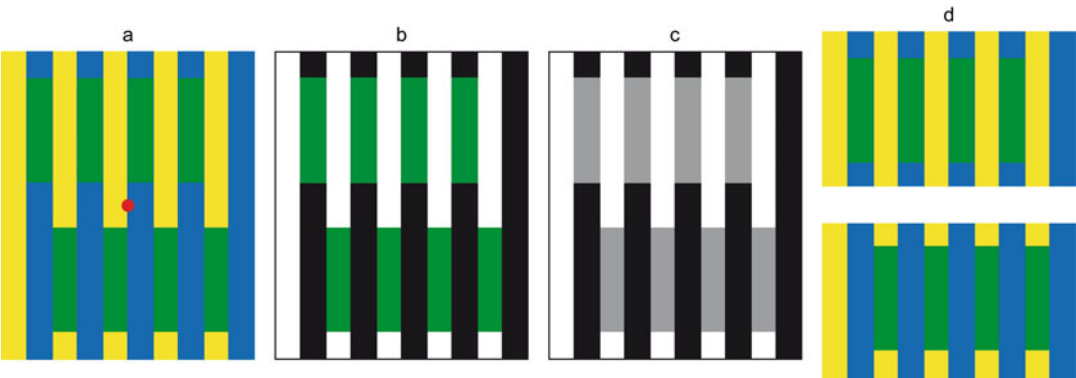
between grouping factors and another relevant Gestalt factor known as *belongingness* [11], which typically supports contrast outcomes in lightness perception. Figure 5 displays an example of the Munker-White illusion [12, 13], which is commonly treated as a contrast illusion governed precisely by belongingness. In Fig. 5c, the grey bars "lying" on the black bars appear lighter than the physically identical grey bars lying on the white ones, thus delivering a result that is similar in direction, but stronger, to simultaneous lightness contrast. Similarly, the green bars on the black ones appear lighter than the green bars on the white ones in Fig. 5b.

If we consider belongingness the major factor responsible for such different appearances in lightness, then one can argue that while in Fig. 4 assimilation is at play, in Fig. 5a contrast is at play despite a certain similarity between the two-color patterns. In Fig. 5a, the top green bars appear yellowish because they contrast with the blue bars on top of which they appear to rest, yet they should be influenced by the yellow bars as they have longer edges in common with the last. In fact, one might still want to argue in favor of assimilation, by maintaining that the yellowish green seen in Fig. 5a is due to an assimilation process determined by the adjacent yellow bars. It is a possibility that, however, clashes with the perceived depth structure of the configuration. When one's attention passes from the top yellowish green bars to the darker and somewhat bluish



Perceptual Grouping and Color, Fig. 4 Colored zebras chromatically modified from one of Vasarely’s graphic artworks. (a) The green stripes appear yellowish on the yellow-striped zebra, while they appear bluish on the blue striped zebra. (b) The red stripes appear with an orange tinge on the yellow-striped zebra, and slightly purplish on

the blue-striped zebra. Moreover, the assimilation effect works both ways: the yellow and blue stripes in (a) appear with a green tinge, while the physically identical yellow and blue stripes in (b) appear to be affected by the red stripes



Perceptual Grouping and Color, Fig. 5 Examples of the Munker-White illusion [12, 13]. In (a) the green bars on top usually appear as lighter and yellowish than the green bars below, which appear darker and somewhat bluish. The configuration is ambiguous as the figure-ground status of the yellow and blue bars change depending on where one fixates. By fixating on the red dot, it is possible also to see one of the two groups of green bars unify to form a semitransparent rectangle. In (b) the effect of contrast is

evident as the lightness of the two sets of green bars change depending on their belongingness to either the black or the white bars. (c) It shows a classic White’s illusion, traditionally treated as a contrast effect. In (d), having split the figure into two, the spatial ambiguity that characterized (a) is no longer present: in the top configuration, the blue bars unify amodally behind the yellow ones to form a blue background on top of which a yellowish green rectangle is seen; in the configuration below the opposite is seen

green bars below, then reorganization of the visual field is experienced due to a process known as *amodal completion* [14]: the visual system groups the two sets of green bars to form two green horizontal rectangles placed one on a uniform blue background, the other on a uniform yellow

background. The bars that do not unify into a continuous surface appear as occluders standing in front of the background and its green rectangle. Because of such grouping processes, Fig. 5a results in a sort of *impossible figure*, with yellow and blue bars changing their figure-ground status

as one inspects the configuration from top to bottom and vice versa. The visual system may attempt to overrun such figure-ground ambiguity by generating a feeble transparency effect, for example, by grouping the top green bars to form a semitransparent rectangle (focusing on the red dot may help see such visual solution). However, this semitransparency, when perceived, does not change the tinge of the green under observation. The shift from the role of figure to that of ground of the yellow and blue bars is instead clearly evident in Fig. 5d, stemming the roles of both grouping principles and belongingness on the appearance of the green bars.

Cross-References

- ▶ [Assimilation](#)
- ▶ [Color Constancy](#)
- ▶ [Color Contrast](#)
- ▶ [Color Phenomenology](#)
- ▶ [Simultaneous Color Contrast](#)

References

1. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt. II. (Investigations on the theory of shape). *Psychol. Forsch.* **4**, 301–350 (1923)
2. Palmer, E.S.: *Vision Science. Photons to Phenomenology*. The MIT Press, Cambridge, MA (1999)
3. Köhler, W.: *Gestalt Psychology*. Liveright, New York (1929)
4. Koffka, K.: *Principles of Gestalt Psychology*. Harcourt, Brace and Company, New York (1935)
5. Fuchs, W.: Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluss von Gestalten (Experimental investigations on the alteration of color under the influence of Gestalten). *Z. Psychol.* **92**, 249–325 (1923)
6. Metzger, W.: *Gesetze des Sehens (The Laws of Seeing)*. Verlag Waldemar Kramer, Frankfurt am Main (1975)
7. Van Lier, R., Wagemans, J.: Perceptual grouping measured by color assimilation: regularity versus proximity. *Acta Psychol.* **97**, 37–70 (1997)
8. Katz, D.: *The World of Colour*. Kegan Paul, Trench, Trubner & Co., London (1935)
9. Beck, J.: *Surface Color Perception*. Cornell University Press, Ithaca/London (1972)
10. Gilchrist, A.: *Seeing Black and White*. Oxford University Press, New York (2006)
11. Economou, E., Zdravkovic, S., Gilchrist, A.: Anchoring versus spatial filtering accounts of simultaneous lightness contrast. *J. Vis.* **7**(12), 2, 1–15 (2007)
12. Munker, H.: *Farbige Gitter, Abbildung auf der Netzhaut und übertragungstheoretische Beschreibung der Farbwahrnehmung [Colored Bars, the Figure on the Retina and Description of Transmission Theory for Color Vision]*. Ludwig-Maximilians-University of Munich, Munich (1970)
13. White, M.: The effect of the nature of the surround on the perceived lightness of grey bars within square-wave test gratings. *Perception*. **10**, 215–230 (1981)
14. Kanizsa, G.: *Organization in Vision*. Praeger, New York (1979)

Philosophy of Color

Barry Maund

Department of Philosophy, The University of Western Australia, Crawley, WA, Australia

Definition

The philosophy of any live topic is a matter of puzzles and problems and debates about their solution. Philosophical debates about color go back to the ancient Greeks, e.g., to Plato, Aristotle, the early atomist Democritus, and many others, and they are still alive. One of the most important reasons why colors are of philosophical interest is that they raise serious metaphysical issues, concerning the nature both of physical reality and of the mind. Central among these issues are questions concerning whether color is part of a mind-independent reality or not and what account can be given of experiences of color.

The Scientific Tradition

Serious thinking about color raises important philosophical puzzles and questions. One of the major problems with color has to do with fitting what people seem to know about colors into what science, particularly physics, tells about physical bodies and their qualities. It is this problem that

historically has led the major physicists, who have thought about color, to hold the view that physical objects do not actually have the colors people ordinarily and naturally take objects to possess. Oceans and skies are not blue in the way that people naively think, nor are apples red (nor green). Colors of that kind, it would seem, have no place in the physical account of the world that has developed from the sixteenth century to the twenty-first century.

Not only does the scientific mainstream tradition conflict with the commonsense understanding of color in this way, but as well the scientific tradition presents a very counterintuitive conception of color. There is, to illustrate, the celebrated remark by the eighteenth-century philosopher David Hume: “Sounds, colors, heat and cold, according to modern philosophy are not qualities in objects, but perceptions in the mind” [1]. Physicists who have subscribed to this doctrine include the luminaries: Galileo, Boyle, Descartes, Newton, Young, Maxwell, and Helmholtz. Maxwell, for example, wrote: “It seems almost a truism to say that color is a sensation; and yet Young, by honestly recognizing this elementary truth, established the first consistent theory of color” [2].

This combination of color eliminativism – the view that physical objects do not have colors, at least in a crucial sense – and color subjectivism – the view that color is a subjective quality – is not merely of historical interest. It is held by many contemporary experts and authorities on color. S. E. Palmer, a leading psychologist and cognitive scientist, writes:

People universally believe that objects look colored because they are colored, just as we experience them. The sky looks blue because it is blue, grass looks green because it is green, and blood looks red because it is red. As surprising as it may seem, these beliefs are fundamentally mistaken. Neither objects nor lights are actually ‘colored’ in anything like the way we experience them. Rather, color is a psychological property of our visual experiences when we look at objects and lights, not a physical property of those objects or lights. The colors we see are based on physical properties of objects and lights that cause us to see them as colored, to be sure, but these physical properties are different in important ways from the colors we perceive. [3]

The examples of authorities who say similar things can be multiplied, e.g., W. D. Wright, E. H. Land, S. Zeki, and R. G. Kuehni. It should be noted, in passing, that there is considerable variation in how the subjectivist component is expressed, with colors being variously described as sensations, psychological properties of visual experiences, mental properties, representations, constructions of the brain, and properties of the brain.

The Complexities of the Scientific Tradition

The traditional scientific thinking on color is more complex, however, than has been presented so far. Central figures such as Descartes, Boyle, and Locke think that there are no colors in the physical world – no colors, as people ordinarily and naively understand them to be. But they also speak of colors as *secondary qualities*, i.e., as powers or dispositions, in bodies, to cause experiences of a certain characteristic type. It is instructive to try to understand this dual position.

It was common for such theorists to make a distinction between colors as they are in physical bodies and colors as they are in sensation. And then, with respect to the first kind of color, they made a distinction between a philosophical way of thinking and the common, everyday way of thinking of such qualities (which Newton describes as “such conceptions as vulgar people would be apt to frame”). As Descartes explains, the ordinary way involves the mistake of “judging that the feature of objects that we call ‘color’ is something ‘just like the color in our sensation’” or of believing that “the self-same whiteness or greenness which I perceive through my senses is present in the body.” The philosophical way of thinking, by contrast, is illustrated in his following remark: “It is clear then that when we say we perceive colors in objects, *it is really just the same as saying* that we perceived in objects something as to whose nature we are ignorant but which produces in us a very clear and vivid sensation, what we call the sensation of color” [4].

So, there is a way – a philosophical way – of thinking of objects as having colors. However, three things are important: (1) it is different from the ordinary way of thinking; (2) it involves a revision, or a reconstruction, of people’s ordinary thinking about color; and (3) it presupposes a notion of “sensation of color” or “color as it is in sensation.” It is crucial how this last notion is understood.

It should also be noted, however, that such theorists as Descartes, Boyle, and Locke sometimes wrote, as well, of colors in physical bodies as those features (“textures”) which grounded the associated powers/dispositions to induce the relevant experiences (sensations). Some commentators think that, in so doing, these theorists were either inconsistent or muddled or found it difficult to make up their minds. However, if Descartes, Boyle, Locke, etc., are in the business of revising or reconstructing ordinary concepts, there is no particular reason why they should not introduce several new concepts to replace the single old one, each one serving a different purpose.

Indeed, it is instructive to look at the famous passage in which Newton writes of the rays not being colored:

For as Sound in a bell, or musical String or other sounding body, is nothing but a trembling motion, and in the Air, nothing but that Motion propagated from the Object, and in the Sensorium ‘tis a sense of that Motion under the form of sound; so Colors in the Object are nothing but a Disposition to reflect this or that sort of Rays more copiously than the rest; in the Rays they are nothing there but Dispositions to propagate this or that motion into the Sensorium, and in the Sensorium they are Sensations of those Motions under the Forms of Colors. [5]

Here, he writes of the colors in the object, colors in the rays, and colors in the sensorium, by analogy with sounds in the object, sounds in the air, and sounds in the sensorium.

There seems to be a similar spirit that inspires modern theorists such as Leo M. Hurvich and D. L. MacAdam. These are color scientists who have stressed the existence of a variety of senses of color terms. Hurvich begins the second chapter of his *Color Vision* with the question “What is color?” which he turns into a series of questions:

Is color something that inheres in objects themselves? Is it related to the light falling on an object? Is it a photochemical event that occurs in the receptor layer of the eye? Is it a neural brain-excitation process? Is it a psychical event? Hurvich’s answer is “Color is all these things” [6]. MacAdam writes:

The term color is commonly used in three distinctly different senses. The chemist employs it as a generic term for dyes, pigments and similar materials. The physicist, on the other hand, uses the term to refer to certain phenomena in the field of optics... Physiologists and psychologists often employ the term in still another sense. They are interested primarily in understanding the nature of the visual process and use the term, on occasion, to denote sensation in the consciousness of a human observer. [7]

One might wonder what unites these various senses of the color terms. Why, that is, should color terms be used to apply to each of these types of conditions or properties – the same terms for each? A plausible answer to this question is the following. If one is interested in the detailed causal mechanisms that underlie color perception, then there is good reason to be interested in different stages of the causal process in which objects cause experience of color: chemists and painters would be interested in the pigments, lighting engineers and physicists would be interested in the light-modifying features of objects, physiologists and psychologists would be interested in the cones and rods in the retinas, other physiologists and neuroscientists would be interested in other neural processes and pathways, and psychologists and philosophers would be interested in the experience of color and in how colors appear. What unites them all is that the theorist is interested, at each stage, in that feature of the stage that contributes to the *perception* of color, i.e., to the way colors *appear*.

Philosophical Responses

Within the philosophical community, there has been a mixed response to the view expressed in this scientific tradition. While some philosophers share key aspects of the view, at least in broad

terms and with certain modifications, there have been many who have rejected the view. There has been strong resistance from many philosophers, both to the eliminativist tendency within the scientific tradition and to the related subjectivism.

One form this resistance takes reflects the fact that each component of this traditional view is very puzzling. One response is to say that color terms – red, blue, purple, orange, yellow, green, brown, etc. – are in order; there are paradigms of colors to which the color terms apply: ripe lemons are yellow, tomatoes and rubies are red, and so on. There is no trouble, by and large, in learning these terms and teaching them in ostensive practices to children and others. Moreover, colors play a central role in so many of human social practices, such as painting, art, fashion, decoration, ceremonial, enjoying nature, etc. In the second place, it is hard to make sense of the claim that colors are properties of sensations or are psychological properties; if they are anything, they are, surely, properties of objects and light sources – of peaches and emeralds, of skies, of rainbows, of glasses of wine, of headlamps, and so on.

This second issue is of key importance. Many color scientists are agreed in thinking that one role for color terms is in application to sensations. As well as the examples already given, it is possible to find, in authoritative texts, definitions like “Color attributes are *attributes of visual sensations*, e.g., hue, saturation and brightness”; “Hue: *attribute of color perception* denoted by the terms yellow, red, blue, green and so forth”; “Brightness is the *attribute of a visual sensation* according to which a given visual stimulus appears to be more or less intense”; etc. However, it is a serious philosophical question as to what is meant by “attributes of sensation” (and likewise “colors in sensation”). There are, for example, different ways of understanding what they are meant to be:

1. Qualities *of* the sensation
2. Qualities presented *in* sensation (or experience)
3. Qualities *represented by* the sensation

Each of these views has had, and continues to have, its champions. Some philosophers have

argued that the first candidate makes no sense, that it commits something like a category mistake. It is interesting that one of the first to argue this way was not a philosopher, but the famous color scientist Ewald Hering [8]. Not surprisingly, perhaps, there are other philosophers who reject this charge. One way to do so is to argue that the perceiver is not aware of the qualities *as qualities of sensations*, even though that is what they are.

One issue that complicates the debates is that psychologists have become increasingly reluctant to use the term “sensation.” For example, modern psychological texts, even ones with titles such as *Sensation and Perception*, tend not to use the term “sensation,” opting for “perception” instead. Indeed, in many of the situations in which traditional color theorists would have used the term “sensation of color,” modern theorists would use such terms as “perceived color” and “apparent color.” Interestingly, examples of both usages can also be found in Newton’s theory. In the famous passage in which he declares that the rays are not colored, “in them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Color,” he begins by saying: “The homogeneous light and rays which *appear red*, or rather make objects *appear* so, I call red-making; those which make objects *appear* yellow, green, blue and violet, I call yellow-making, green-making, blue-making, violet-making, and similarly for the rest.”

Accordingly, instead of using the expressions “attribute of sensation” and “sensation of color,” it would make sense to use “apparent color” and “perceived color,” as many theorists currently do, or “color as we see it,” as some philosophers suggest. So, the debates can be framed in terms of the colors objects *appear to have* or *look to have*. The philosophical issues remain, however. The point is that it is a matter of philosophical dispute as to what exactly it is for something to *look blue*, say, (i.e., to *appear blue*). It is plausible that it involves the object causing an experience in the subject, but then the question can be raised with respect to the report, say, about one’s experience of blue, whether that report is describing:

1. A quality *of* the experience
2. A quality which is *presented in* the experience
3. A quality which the experience *represents* the object as having

Not only that, but, supposing that the experience does have representational content, there is a further question as to whether it has other features as well (ones that are open to introspective access). One view is that there are sensory qualities which are presented in sensory experience and which are “projected” onto physical objects. In this way, objects appear to have qualities that they do not actually have. Since this “projection” is not a literal one, it requires explanation of course, but its defenders think that it can be provided. On this theory, the sensory qualities presented in experience also form part of the representational content of the experience. This view is sometimes combined with *color fictionalism*, the view that ordinary color language is a language about properties which objects appear to have (but do not actually have). Objects in the world are such that “it is as if they have the colors.”

Color Realism

These debates raise questions about the strength of the argument against the reality of color. Consider the passage from S. E. Palmer, quoted in the first section. It is possible to see it as presenting an argument against the reality of colors. It takes the form: when objects are seen as colored, the objects are experienced in certain ways, they are seen as having certain qualities, but the objects do not have those qualities; the colors perceived are different from any the physical objects possess. There are two claims implicit here: (i) the colors objects are perceived as having – the “apparent colors” – a certain distinctive character, and (ii) the physical sciences have shown that no qualities with that particular character play any causal part in the perception of colors. From this, it is concluded that neither objects nor lights are colored in anything like the way they are experienced.

This argument depends on one’s being able to characterize the distinctive character of the “apparent colors.” Here is an important point on which philosophers divide. One group holds that it is obvious that this distinctive character includes a qualitative aspect, one that is familiar to common perceivers. It is by virtue of having this qualitative character that the various colors fit together collectively, in system of relationships of similarities and differences (a psychological color space). In this system, for example, orange is between red and yellow, resembling each, whereas it is not between blue and green, nor blue and red. Another group of philosophers, however, holds that any such qualitative aspect is no part of colors themselves, where the colors are the properties one’s experiences *represent* objects as having. (It may be part of the way colors are represented, rather than essential to the color itself. Think here of the difference between *heat* in a body and *the way things are experienced* as hot or cold.) The claim may be put as follows: (1) colors are certain properties, ones that objects look to have (appear to have) when one sees the objects; (2) the properties objects look to have are the qualities one’s experiences *represent* the objects as having.

Everything now turns on what theory of *representation* one can call upon. Once again opinions (“carefully considered positions”) differ. On one view, the qualities are ones that play the appropriate causal role in the production of the experiences of color. As a consequence, so it is argued, these qualities are ones that it requires scientific inquiry to locate. Accordingly, it is thought there are good reasons to think that surface colors of objects, for example, are spectral reflectances. On the other hand, another group of philosophers reject that theory of representation, arguing that the qualities represented are dispositional properties: dispositions to cause experiences of the right type.

Then again, there are other philosophers who defend a different version of color realism. They accept that the colors do have the relevant qualitative character, but deny that the scientists provide good argument against the reality of color. That is to say, they accept that the ordinary conception of color is of features with a simple,

qualitative character, one that is typically revealed in color perception, but they deny that the scientific argument against the reality of such qualities is cogent. Contemporary defenders of this view appeal to the claim that colors are qualities that are supervenient on light-modifying features of physical bodies. The light-modifying features cause visual experiences, but, so it is argued, colors are qualities that are revealed in experience.

There is another argument against color realism, which has a long pedigree, but which has come to prominence in recent times. One of its strongest advocates has been C. L. Hardin [9]. Any defense of color realism has to acknowledge and defend a distinction between *real color* and (merely) *apparent color* or between *real color* and *illusory color*. Any such distinction, however, needs to rely on a distinction between standard conditions and nonstandard ones and between normal, or standard, observers and nonstandard observers. Things which are *really blue*, say, are ones that will look blue to standard perceivers, in standard conditions. In other sorts of conditions, or to other kinds of perceivers, they may look to have other colors. If so, those other colors will be illusory ones, or “merely apparent” ones.

Hardin has argued that there is no way of identifying such conditions, and such perceivers, that is not arbitrary. A uniformly colored square, for example, will appear differently if the sides are, respectively, 2, 10, and 100 cm. And they will look to have different colors, when set against different backgrounds. The problem is even more pronounced in the case of standard observers. Competent human color perceivers divide into a number of groups that differ, for example, on which hues are instances of the unique hues: unique blue, red, green, and yellow. The differences are quite dramatic. Not only that, but there are significant differences between males and females, with respect to these abilities.

This kind of problem has led some theorists, e.g., Jonathan Cohen, to defend another form of color realism, a version of color relationalism [10]. On this theory, color is a relational property, defined with respect to relations between the object, the environmental viewing conditions, and the kind of perceiver. On this view, there are

no such properties as *blue simpliciter*, *red simpliciter*, etc.; there is only blue-for-perceiver kind A, in conditions C; blue-for-perceiver kind B, in conditions C*; and so on. It should be noted that although Cohen calls the account “realist,” it is a special type of realism. On this account, a whole host of relational properties are equally real.

Color Relationalism

There are other forms of color relationalism which have links to action-based theories of perception, as developed principally by the psychologist J. J. Gibson. A leading example is the theory defended by Evan Thompson, the *Ecological View of Colors*. On this account, colors are taken to be dependent, in part, on the perceiver and so are not intrinsic properties of a perceiver-independent world. Being colored, instead, is construed as a relational property of the environment, connecting the environment with the perceiving animal. In the case of the color of physical surfaces, “being colored corresponds to the surface spectral reflectance as visually perceived by the animal” [11]. Thompson’s work has been supported by that of Alva Noe, who defends what he calls “an enactive approach to perception” [12]. As in the case of Thompson, his theory is guided by Gibson’s work, albeit with modification.

More recently, other forms of relational theories have emerged. Of these one of the most important is the version defended by M. Chirimuuta, in her book, *Outside Color*, 2015 [13]. Chirimuuta advocates a version of color relationalism which she calls *Color Adverbialism*: On this account, colors are not properties of external physical objects, or of brains, or of our mental states; instead, they are properties of perceptual processes or interactions which involve objects, brains, and mental states. So, according to color adverbialism, “seeing the colors” means “seeing in a colourful way.” She appeals to recent research on color vision, relied to the function of color vision, to argue that it provides great difficulties for objectivist color realism – both primitivist and reductionist physicalist. While she admits that anti-realism, especially the

form favored by Hardin, is in a better position, they think that color adverbialism presents the best explanation.

Another important recent contributions to the philosophy of color is the book by Matthen [14]. In this work, Matthen articulates a theory of sense-perception in which color plays a prominent role. The work is significant for the theory of color that he presents, one that draws heavily on comparative studies of color vision among different species. Matthen replaces his earlier objectivist views on color by an account that has more in common with the ecological theory favored by Thompson.

Goethe's *Theory of Colors*

This entry would not be complete without reference to the color theory of the famous German author Johann Wolfgang von Goethe. Goethe was deeply critical of Newton's analytical approach and of his theoretical explanations of color phenomena [15]. While mainstream color science has been mostly dismissive of Goethe's attack, there are thinkers who find a positive side to Goethe's work.

There are two aspects to Goethe's critique. One is that he describes experiments that, he claims, contradict Newton's theory. The second aspect is that he provides a set of observations that Newton's theory does not, and cannot, explain. Additionally, he claims to provide an alternative explanation for these phenomena. H. Helmholtz, the nineteenth-century physicist and color scientist, rejected the criticisms of Newton's science as based on serious misconceptions [16]. First of all, Newton was well aware of some of the experiments described, and, more importantly, all of these experiments can be satisfactorily handled by Newton's theoretical approach. With respect to the other observations, it is true that Newton does not explain them, but these phenomena are not ones that his theory is designed to explain. Nevertheless, Helmholtz praised Goethe's descriptions, of both the experiments and the other observations, as accurate and important.

The issues are perhaps best brought out by a response to Goethe's work, by the philosopher L. Wittgenstein [17]. For a clear exposition, see an article by another philosopher, Zeno Vendler [18]. Wittgenstein points out that Goethe, despite what he thinks, does not provide a scientific theory at all; his activity has a different nature. What he is doing is better described as providing phenomenological observations, i.e., observations of how colored objects appear and of the relations between those appearances, e.g., that yellow (of the fullest saturation) is lighter than red. The resulting propositions that emerge from this sort of activity are not ones that Newton's theory is designed to address, but they are very important, nevertheless. (Whether they can be explained by Newton's theory supplemented by neuroscience is not so easily decided.)

Cross-References

- [Appearance](#)
- [Color Order Systems](#)
- [Color Phenomenology](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Hering, Karl Ewald Konstantin](#)
- [Psychologically Pure Colors](#)
- [Subjective Colors](#)
- [Unique Hues](#)

References

1. Hume, D.: Treatise of human nature. In: Lindsay, A.D. (ed.) Dent, London (1738/1911), book III, part I, sect. 1, p. 177; book I, IV, IV, p. 216
2. Maxwell, J.C.: On color vision. In: MacAdam, D.L. (ed.) Sources of Color Science, p. 75. MIT Press, Cambridge, MA (1890/1970)
3. Palmer, S.E.: Vision Science, p. 95. MIT Press, Cambridge, MA (1999)
4. Descartes R.: Principles of philosophy. In: Cottingham, J., et al. (eds.) Descartes: Selected Writings (1644/1988), par. 70; see also par. pp. 68–70
5. Newton, I.: Optics. In: MacAdam, D.L. (ed.) Sources of Color Science, pp. 23–24. MIT Press, Cambridge, MA (1730/1970)
6. Hurvich, L.M.: Color Vision, p. 2. Sinauer, Sunderland (1981)

7. MacAdam, D.L.: *Color Measurement: Themes and Variations*, p. 1. Springer, Berlin (1985)
8. Hering, E.: *Outline of a Theory of the Light Sense* (trans. Hurvich, L.M., Jameson, D.). Harvard University Press, Cambridge, MA (1920/1964)
9. Hardin, C.L.: *Color for Philosophers*. Hackett, Indianapolis (1988/1993)
10. Cohen, J.: *The Red and the Real*. Oxford University Press, Oxford, UK (2009)
11. Thompson, E.: *Color Vision*, Chapter 5, Routledge, London (1995)
12. Noë, A.: *Action in Perception*. MIT Press, Cambridge, MA (2004)
13. Chirimuuta, M.: *Outside Color: Perceptual Science and the Puzzle of Color*. MIT Press, Cambridge, MA (2015)
14. Matthen, M.: *Seeing, Doing and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press, Oxford, UK (2005)
15. Goethe, J.W.: *Goethe's Theory of Colors* (trans. Eastlake, C.L.). Cassell, London (1840/1967)
16. Helmholtz, H.: *Goethe's scientific researches*. In: Helmholtz, H. (ed.) *Popular Scientific Lectures*, pp. 1–21. Dover, New York (1881/1962)
17. Wittgenstein, L.: In: Anscombe, G.E.M. (ed.) *Remarks on Color*. Blackwell, London (1977)
18. Vendler, Z.: Goethe, Wittgenstein, and the essence of color. *Monist*. **78**, 391–418 (1995)

Phosphors and Fluorescent Powders

Weidong Zhuang and Jiyong Zhong
National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Synonyms

[Luminescent materials](#)

Definition

Phosphor (fluorescent powder) is a kind of inorganic materials which absorbs the energy from electrons, photons, or even other forms of microscopic particles and converts it into visible light. Phosphor (fluorescent powder) has a history of

about 100 years and today plays one of the key roles in lighting, display, and imaging. So far there are varieties of phosphors finding their applications in cathode-ray tubes (CRTs), fluorescent lamps, x-ray films, plasma display panels (PDPs), and recently developed white light-emitting diodes (w-LEDs).

The Mechanism of Luminescence

Although phosphors are available in diversity, the mechanisms of luminescence in them are similar, including the following primary items:

- Transitions of rare-earth ions and transition metal ions
- Crystal field splitting and nephelauxetic effect
- Energy transfer

Transitions of Rare-Earth Ions and Transition Metal Ions

Essentially, luminescence of phosphors is a process of electron transitions, which occurs in luminescence center ions (active centers). But not all of the ions can be considered to act as optically active centers, which are usually based on rare-earth ions or transition metal ions. In this section, we will concentrate on these luminescent centers and discuss their energy levels.

Electronic Configuration, Energy Levels, and Transitions of Rare-Earth Ions

The rare-earth ions usually refer to lanthanide ions, Sc^{3+} and Y^{3+} . The electronic configurations of both trivalent and divalent rare-earth ions in the ground states are shown in Table 1. It can be seen that the 4f orbit is empty for Sc^{3+} , Y^{3+} , and La^{3+} and fully filled for Lu^{3+} , which will lead to no luminescence due to lack of energy levels generated by unpaired 4f electrons, while the lanthanide ions from Ce^{3+} to Yb^{3+} have partially filled 4f orbitals, thus they exhibit their characteristic energy levels, which can induce luminescence processes. Additionally, there are three divalent lanthanide ions, Sm^{2+} , Eu^{2+} , and Yb^{2+} , and their electronic configurations are the same as those of

Phosphors and Fluorescent Powders, Table 1 Electronic configurations and ground state of rare-earth ions

Ions	4f electrons	Ground state	Ions	4f electrons	Ground state
Sc ³⁺	0	¹ S ₀			
Y ³⁺	0	¹ S ₀			
La ³⁺	0	¹ S ₀			
Ce ³⁺	1	² F _{5/2}			
Pr ³⁺	2	³ H ₄			
Nd ³⁺	3	⁴ I _{9/2}			
Pm ³⁺	4	⁵ I ₄			
Sm ³⁺	5	⁶ H _{5/2}	Sm ²⁺	6	⁷ F ₀
Eu ³⁺	6	⁷ F ₀	Eu ²⁺	7	⁸ S _{7/2}
Gd ³⁺	7	⁸ S _{7/2}			
Tb ³⁺	8	⁷ F ₆			
Dy ³⁺	9	⁶ H _{15/2}			
Ho ³⁺	10	⁵ I ₈			
Er ³⁺	11	⁴ I _{15/2}			
Tm ³⁺	12	³ H ₆			
Yb ³⁺	13	² F _{7/2}	Yb ²⁺	14	¹ S ₀
Lu ³⁺	14	¹ S ₀			

Eu³⁺, Gd³⁺, and Lu³⁺, respectively. The electronic state is denoted as ^{2S+1}L_J, where S, L, and J are the spin angular momentum, orbital angular momentum, and total angular momentum, respectively.

The experimentally measured energy levels of trivalent lanthanide ions in lanthanum fluoride (LaF₃) are illustrated in *Dieke diagram* (as shown in Fig. 1). It presents the energy of electronic states ^{2S+1}L_J for trivalent ions in LaF₃ and almost any other crystals. The number of levels is determined by the ion species and degree of energy level degeneracy, which can be influenced by the symmetry of the crystal field surrounding the lanthanide ions, and the width of each level is a reflection of the crystal field splitting. By taking advantage of the *Dieke diagram*, one can explain the absorption and emission spectra and identify each spectral peak of rare-earth ions in different luminescent materials. Especially, the luminescence of trivalent rare-earth ions with f-f transitions can be assigned according to *Dieke diagram*. An other group of rare-earth ions, represented by Ce³⁺ and Eu²⁺ with f-d transitions, deserve much more attentions for their higher luminescent efficiency, which will be discussed as following.

Ce³⁺: The Ce³⁺ ion has the simplest electron configuration among the rare-earth ions. The 4f¹ ground-state configuration is divided into two sublevels, ²F_{5/2} and ²F_{7/2}, and these two sublevels are separated by about 2,000 cm⁻¹ as a result of spin-orbit coupling. This is the reason for the double bands usually observed in the Ce³⁺ emission spectra. The 5d¹ excited state configuration is split into two to five components by the crystal field, with the splitting number depending on the crystal field symmetry. The Ce³⁺ emission is strongly affected by the host lattice through the crystal field splitting of the 5d orbital and the nephelauxetic effect and usually covers the visible light region.

Eu²⁺: The Eu²⁺ ion has the ground state of 4f⁷ (⁸S_{7/2}) and the excited state of 4f⁶5d¹. The luminescence is strongly dependent on the hosts, with the emission color covering ultra-violet to red. The abundant emission colors are due to the different covalency (nephelauxetic effect) and crystal field splitting from one host to another. With large covalency or crystal field splitting, the 5d energy level of Eu²⁺ is lowered greatly, leading to the red shift of the absorption and emission bands.

Transitions of Transition Metal Ions

Transition metal ions are an important class of dopants in luminescent materials and solid-state laser materials, and their emission colors are very abundant. Among the transition metal ions, Mn²⁺ and Mn⁴⁺ are two of the most important and useful luminescent centers in phosphors, which will be discussed below.

Mn²⁺: The emission color of Mn²⁺ always exhibits green or red, depending on the crystal field strength of the host crystal. The emission spectrum of Mn²⁺ displays a broad band, and the emission corresponds to the ⁴T₁ (⁴G) → ⁶A₁ (⁶S) transition. In general, the emission color is green when Mn²⁺ is tetrahedrally coordinated, and it is red when Mn²⁺ is octahedrally coordinated. The absorption of Mn²⁺ is usually low due to the spin-forbidden transition. To improve the absorption of Mn²⁺, energy

This effect is called crystal field effect, which has shown great influence on luminescence ions. Crystal field splitting is, in fact, the energy level splitting of luminescence ions by the influence of crystal field. Generally, the crystal field splitting is affected by symmetry of coordination anions and the distances between luminescence ions and coordination anions. And crystal field splitting is usually described by the point charge electrostatic model (PCEM), which can be usefully applied especially in the case of coordination in the form of an octahedron, cube, tricapped trigonal prism, and cuboctahedron.

Some electrons of the ligands move into the orbitals of the central ion and reduce the cationic valency. Due to this reduction, the d-electron wave functions expand toward the ligands to increase the distances between electrons, reducing the interaction between them. This effect is called the nephelauxetic effect. This effect may be considered to increase with ligands anion in the order: $F^- < O^{2-} < N^{3-} < Cl^- < Br^- < I^- < S^{2-}$.

Energy Transfer

Energy transfer means energy migrating from a higher energy ion to another lower energy ion or even from a host crystal to activators. Thus, three kinds of energy transfer should be included as following:

- *The energy transfer between the same ion species.* When the ionic separation becomes smaller, the energy transfer probabilities increases by means of resonant, exchange interaction, and multipolar interaction. As a result, the optical exciton is trapped at defects or impurity sites, enhancing non-radiative relaxation, which is the reason for concentration quenching.
- *The energy transfer between different ion species* can take place when they have closely matched energy levels. The energy transfer result is that the luminescence of energy-accepted activators enhanced and the luminescence of energy-donated activators weakened. Typically, Ce^{3+} ions transfer energy to Tb^{3+} ions in the phosphors $CeMgAl_{11}O_{19}:Ce^{3+}$, Tb^{3+} ; $LaPO_4:Ce^{3+}, Tb^{3+}$; etc.

- *The energy transfer from a host crystal to activators* mainly refers to energy transfer from an excited oxy-anion complex to lanthanide ions. This kind of energy transfer is commonly seen in vanadate, molybdate, tungstate, etc., such as $YVO_4:Eu^{3+}$, $CaMoO_4:Eu^{3+}$, and $Y_2WO_6:Eu^{3+}$.

Characterization

Besides the mechanism of luminescence, characterization of the prepared phosphors is vital to evaluate the practical application. The principal characterization of phosphors can be listed as the following:

- Spectrum
- Quantum efficiency
- Thermal quenching
- Fluorescent lifetime

Spectrum

When light interacts with a phosphor, refraction, absorption, scattering, and luminescence will occur. The luminescence process – excitation and emission – will be discussed and characterized by excitation and emission spectra.

Emission Spectrum

During the luminescence process, the distribution of emission photon number varies with emission wavelength or frequency, that is, the so-called emission spectrum, which is important to judge what color of light is emitted by the phosphor and to calculate the color coordinates. Generally, two kinds of emission spectrum are presented: line spectrum for activators with f-f transitions and band spectrum for activators with d-electron transitions. No matter what kinds of emission spectrum, it shows difference with activator species and compositions of host.

Excitation Spectrum

The distribution of emission photon number of the monitored wavelength varies with excitation wavelength or frequency. Thus, it, in fact, is a specific emission spectrum, which is used to

characterize the effectiveness of excitation. From this spectrum, it is convenient to judge the effective excitation wavelength for phosphors, which is important for practical use.

Quantum Efficiency

The emission and excitation spectra let people know what light the phosphor emits and the effectiveness of different excitation wavelength, while it still does not know how efficient the energy conversion is. Thus, the quantum efficiency is introduced. The quantum efficiency of a phosphor is how much absorbed light (photons) is converted into emission, which is in the range of 0–1. This is also called internal quantum efficiency. The external quantum efficiency denotes the ratio of the number of photons of the emitted light to that of the incident light on a phosphor. The internal (η_i) and external (η_0) quantum efficiency can be computed by the following equations:

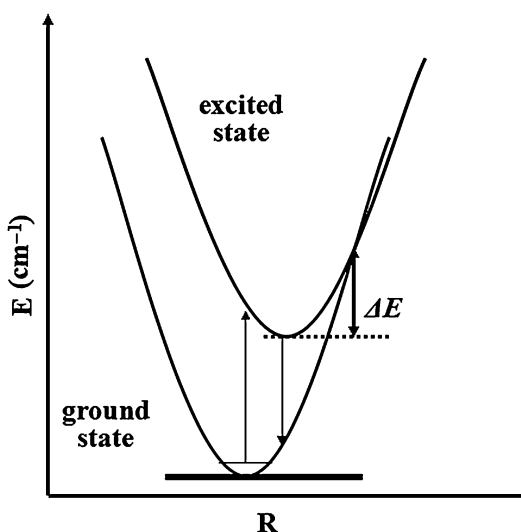
$$\eta_i = \frac{\int \lambda \cdot P(\lambda) d\lambda}{\int \lambda \{E(\lambda) - R(\lambda)\} d\lambda}$$

$$\eta_0 = \frac{\int \lambda \cdot P(\lambda) d\lambda}{\int \lambda \cdot E(\lambda) d\lambda}$$

where $E(\lambda)/h\nu$, $R(\lambda)/h\nu$, and $P(\lambda)/h\nu$ are the number of photons in the spectrum of excitation, reflectance, and emission of the phosphor, respectively.

Thermal Quenching

Thermal quenching of luminescence is the phenomenon in which the luminescence efficiency decreases as the temperature increases due to the increasing phonon–electron interaction. This phenomenon can be explained by configuration coordinate (as shown in Fig. 2). The excited electron can go across the crossing point of ground state and excited state and relax to the ground state non-radiatively, the probabilities of which will increase with the increasing temperature. The



Phosphors and Fluorescent Powders, Fig. 2 Configurational coordinate diagram for thermal quenching

energy difference between the lowest vibrational level of excited state and the crossing point is called activation energy for thermal quenching, which is used to judge the thermal quenching properties of phosphors. The larger value of activation energy means higher energy barrier for thermal quenching. The activation energy can be calculated by the Arrhenius equation:

$$I_T = \frac{I_0}{1 + c \exp\left(\frac{-\Delta E}{kT}\right)}$$

Here I_0 is the initial emission intensity, I_T is the intensity at different temperatures, ΔE is the activation energy of thermal quenching, c is a constant for a certain host, and k is the Boltzmann constant.

Fluorescent Lifetime

Fluorescent lifetime is another important factor, which is not used to judge the phosphors good or bad but limits the applications of phosphors. For display and imaging applications, short fluorescent lifetime is needed as well as high luminescence efficiency and narrow full width at half maximum (FWHM) of emission spectrum. The decay process of the luminescence intensity $I(t)$

after the termination of excitation at $t = 0$ is generally represented by an exponential function of the elapsed time after the excitation:

$$I(t) = I_0 \exp(-t/\alpha)$$

where α is the decay time constant of the emission and $\tau = -1/\alpha$ is named fluorescent lifetime. For convenience, fluorescent lifetime is defined as the duration of $I(t)$ decreasing from I_0 to $1/eI_0$. Generally, light emission due to a forbidden transition has a long fluorescence lifetime, while that due to an allowed transition usually shows a short fluorescence lifetime, but that is not absolute since the electron may be captured by a trap or lattice defect, leading to long after-glow luminescence. Fluorescent lifetime is usually determined by measuring the time-resolved emission spectroscopy, which provides information on crystal structure and the kinetic behavior of luminescent excited states and intermediates as well.

Practical Phosphors

As mentioned above, phosphors are various and have shown varieties of applications in lighting, display, and imaging. However, the lighting industry constitutes the largest consumer of phosphors and produces the largest quantity of phosphor-related products. Thus, the widely used phosphors for fluorescent lamps and w-LEDs are chosen and introduced as following:

- Phosphors for fluorescent lamps
- Phosphors for w-LEDs

Phosphors for Fluorescent Lamps [2]

The invention of fluorescent lamp can be traced back to the nineteenth century while the calcium halophosphate phosphor was developed, and the age of fluorescent lamps was truly entered since 1950s. However, the problems of poor color-rendering and high color temperature of this lamp prompted the development of three-band fluorescent lamp, which shows higher color-rendering and even higher luminescence

efficiency. The practical phosphors for fluorescent lamps will be introduced briefly as the following.

Calcium Halophosphate Phosphors for Traditional Fluorescent Lamps

Crystal structure: These phosphors have an apatite-type structure belonging to the hexagonal system. The Ca^{2+} ions have two different sites: Ca (I) coordinated by six oxygen and Ca (II) coordinated by halogen ions. Sb^{3+} and Mn^{2+} ions are capable of replacing Ca^{2+} ions at both types of sites. In addition, the Ca^{2+} ions are possible to be partially replaced by Sr^{2+} , Ba^{2+} , and Cd^{2+} ions.

Luminescent properties: These series of phosphors are doubly activated by Sb^{3+} and Mn^{2+} . The Sb^{3+} emission peak is located at ~ 480 nm and is not influenced by the kind of halogen in the host. The emission peak of Mn^{2+} is located at 575–585 nm, depending on the F:Cl ratio. Energy transfer from Sb^{3+} to Mn^{2+} can be observed and show great influence on emission colors by turning $\text{Sb}^{3+}:\text{Mn}^{2+}$ ratio. By adjusting appropriate compositions of these phosphors, the color temperatures of the white light can be obtained ranging from 3,000 to 6,500 K. In addition, these series phosphors have almost 90% of the emission intensity retained, even at 200 °C, without a shift in the emission peak.

Phosphors for Three-Band Fluorescent Lamps

Phosphors utilized for three-band lamps can be grouped into three categories as following:

1. Red phosphor: $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ is the only available red phosphor for three-band lamps mainly due to its nearly perfect luminescence properties, except for the high price.

Crystal structure: The crystal structure of this phosphor is cubic, and Eu^{3+} ions occupy the two Y^{3+} sites of C_2 and S_6 symmetry.

Luminescent properties: The excitation spectrum has a broad band ranging from 200 to 300 nm, which matches perfectly with 254 nm light emitted by mercury vapor. The emission spectra contains a very narrow and intense emission band at 611 nm (originates from the $^5\text{D}_0 \rightarrow ^7\text{F}_2$ transition of Eu^{3+}) with a

full width at half maximum of ~ 7 nm. This phosphor exhibits high color purity, high quantum efficiency (nearly 100%), excellent thermal quenching properties, and stable chemical properties.

2. Green phosphors: There are three kinds of practicable green phosphors being developed successively, which are $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$; $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$; and $\text{GdMgB}_5\text{O}_{10}:\text{Ce}^{3+}, \text{Tb}^{3+}$.

Crystal structures: $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ belongs to the hexagonal system and has the same magnetoplumbite structure as the $\text{PbFe}_{12}\text{O}_{19}$ crystal. $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$ belongs to the monazite crystal group. $\text{GdMgB}_5\text{O}_{10}:\text{Ce}^{3+}, \text{Tb}^{3+}$ has monoclinic symmetry and is isostructural with $\text{LaCoB}_5\text{O}_{10}$.

Luminescent properties: The common characteristics of these phosphors are activated by Tb^{3+} and adopting Ce^{3+} as sensitizer, because the absorption lines of Tb^{3+} in the ultraviolet region does not match the light emitted by mercury vapor, but Ce^{3+} can absorb the light emitted by mercury vapor efficiently and transfer the energy to Tb^{3+} with a high efficiency by overlapping the emission spectra of Ce^{3+} and absorption lines of Tb^{3+} . Thus, these phosphors exhibit the emission of Tb^{3+} with emission peak at ~ 543 nm (corresponding to the $^5\text{D}_4 - ^7\text{F}_5$ transition) and a full width at half maximum of ~ 10 nm. But by comparing other properties, the $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ presents much more excellent in quantum efficiency (as high as 97%), thermal quenching properties, and chemical stability; therefore, the $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ phosphor is still one of the most important and widely used green phosphors for three-band fluorescent lamps.

3. Blue phosphors: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ and $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ are two main classes of blue-emitting phosphors.

Crystal structures: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ has an apatite-type structure belonging to the hexagonal system. $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ has a crystal structure similar to hexagonal β -alumina.

Luminescent properties: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ gives a sharp emission spectrum peaking at 447 nm and a broad band

excitation spectrum covering 200–400 nm region. Partial replacement of Sr^{2+} by Ca^{2+} or Ba^{2+} results in red shift or blue shift, mainly ascribed to the change of crystal field. Incorporation of a small amount of Ba^{2+} serves to improve the degradation characteristics; incorporation of a small amount of borate may increase the emission intensity. $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ also has a broad band excitation spectrum in the ultraviolet region and an emission band with the peak located at ~ 450 nm. Generally, the luminescent properties can be adjusted by controlling the Al_2O_3 ratio, Eu^{2+} concentration, and Sr^{2+} replacing part of Ba^{2+} . By comparison, $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ shows higher quantum efficiency (appears to exceed 100% measured by conventional methods). The relative poor thermal quenching properties are still challenges for both blue phosphors.

Phosphors for W-LEDs

W-LEDs are considered to have brought about a revolution in energy-efficient lighting due to the outstanding advantages such as long lifetime, high energy efficiency, and environmental friendliness. Currently, w-LEDs are commonly generated by the combination of the blue LED chips and yellow-emitting phosphors. But, the white light obtained by this approach would result in low color-rendering index and high correlated color temperature due to the red emission deficiency in the visible spectrum. Generally, red-emitting phosphors are added to make up for this deficiency. Thus, the widely used phosphors for w-LEDs can be divided into two groups as the following:

Oxide Yellow Phosphors

1. $\text{YAG}:\text{Ce}^{3+}$ garnet phosphor

Crystal structure: YAG is the abbreviation of $\text{Y}_3\text{Al}_2\text{Al}_3\text{O}_{12}$, which has the classical garnet structure represented by the general formula $A_3B_2X_3\text{O}_{12}$, where A , B , and X are eight, six, and four coordinated with the surrounding O, forming dodecahedron, octahedron, and tetrahedron, respectively. The octahedron and tetrahedron do not share any edge among themselves, but they share edges with at least one dodecahedron.

Luminescent properties: The excitation spectrum of YAG:Ce^{3+} is mainly composed of two broad bands with excitation peaks located at 340 and 460 nm, respectively. The emission spectrum shows a very broad band with emission peak at ~ 535 nm and a full width at half maximum (FWHM) of ~ 100 nm. The high external quantum efficiency (more than 83%), excellent thermal quenching properties (more than 87% of the initial emission intensity remained at 150°C), and the strong absorption of blue light emitted by blue LED chip made YAG:Ce^{3+} one of the most important and widely used phosphors for white LEDs. Actually, YAG:Ce^{3+} is a green-yellow-emitting phosphor, and the emission spectrum in the red region is very weak, which results in the produced white light having a high color temperature. Generally, Gd^{3+} replacing part of Y^{3+} or increasing Ce^{3+} concentration may turn the emission spectrum shift toward longer wavelength, but the quantum efficiency decreases radically; meanwhile, the thermal quenching increases.

2. $\text{Sr}_2\text{SiO}_4\text{:Eu}^{2+}$ silicate phosphor

Orthosilicate phosphor has outstanding advantages such as high luminescence efficiency, low cost, and strong excitation in UV to blue light region, which have attracted lots of attentions.

Crystal structure: Sr_2SiO_4 has two types of crystal structures, which are orthorhombic corresponding to α' phase and monoclinic corresponding to β phase. The α' and β phases of Sr_2SiO_4 are isostructural with Ba_2SiO_4 and $\beta\text{-Ca}_2\text{SiO}_4$, respectively. Generally, in order to stable the $\alpha'\text{-Sr}_2\text{SiO}_4$ phase, Ba^{2+} ions were introduced to replace part of Sr^{2+} ions. For $\alpha'\text{-Sr}_2\text{SiO}_4$, Sr^{2+} ions have two kinds of sites, Sr(1) and Sr(2), coordinated by 10 and 9 oxygen atoms, respectively.

Luminescent properties: The excitation spectrum of $\text{Sr}_2\text{SiO}_4\text{:Eu}^{2+}$ shows a broad band covering the spectral region of 200–550 nm with excitation peak at about 387 nm. The emission spectrum shows two bands centered at about 490 and 560 nm, assigning to the luminescence centers

occupying two types of Sr sites: Sr(1) and Sr(2), respectively. Due to the weak 490 nm emission peak, the emission spectra of $\text{Sr}_2\text{SiO}_4\text{:Eu}^{2+}$ shows emission peak at about 560 nm, which can be shift to longer wavelength by increasing the SiO_2 content. Thermal quenching of this kind of silicate phosphor is unsatisfactory for the emission intensity decreases by 62% when the temperature rises to 150°C compared to the initial intensity measured at room temperature. Though the luminescence efficiency of $\text{Sr}_2\text{SiO}_4\text{:Eu}^{2+}$ is comparable with YAG:Ce^{3+} , the thermal stability and structure stability are still limitations for its applications.

Nitride Red Phosphors [3]

1. $\text{M}_2\text{Si}_5\text{N}_8\text{:Eu}^{2+}$ (M = Ca, Sr, Ba) phosphors

Crystal structure: $\text{Ca}_2\text{Si}_5\text{N}_8$ has a monoclinic crystal system with the space group of Cc, whereas both $\text{Sr}_2\text{Si}_5\text{N}_8$ and $\text{Ba}_2\text{Si}_5\text{N}_8$ belong to orthorhombic crystal system with the space group of Pmn2₁. There are two different crystallographic sites for alkaline earth metals. Each Ca^{2+} ion in $\text{Ca}_2\text{Si}_5\text{N}_8$ is coordinated by seven nitrogen ions, while Sr^{2+} and Ba^{2+} in $\text{Sr}_2\text{Si}_5\text{N}_8$ and $\text{Ba}_2\text{Si}_5\text{N}_8$ are coordinated by eight or nine nitrogen ions.

Luminescent properties: The $\text{M}_2\text{Si}_5\text{N}_8\text{:Eu}^{2+}$ (M = Ca, Sr, Ba) phosphors show very intense red emission under blue light excitation. The emission color changes from orange to deep red depending on the alkaline earth metal ion ratio. The emission spectra all exhibit a broad and asymmetrical band extending from 550 to 850 nm. Upon the 450 nm excitation, $\text{Sr}_2\text{Si}_5\text{N}_8\text{:Eu}^{2+}$ shows an absorption and external quantum efficiency of 87% and 71%, respectively.

2. $\text{CaAlSiN}_3\text{:Eu}^{2+}$ phosphor

Crystal structure: CaAlSiN_3 belongs to orthorhombic crystal system with Cmc2₁ space group. The Ca^{2+} ions have only one crystallographic site, which is coordinated by five nitrogen ions.

Luminescent properties: The excitation spectrum covers a broad spectral range of 250–600 nm, showing an extremely broad band and strong absorption of visible light.

The emission spectrum extends from 550 to 800 nm, showing a broad band centered at 655 nm with a FWHM value of 93 nm. When the phosphor is excited by the blue light ($\lambda = 450$ nm), the external quantum efficiency is up to 70%. In addition, it has a slight thermal quenching. The luminescence is quenched by 10% of the initial intensity when the phosphor is heated to 150 °C and excited at 450 nm.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Light, Electromagnetic Spectrum](#)
- [Luminescence](#)
- [Quantum Dots](#)

References

1. Carnall, W.T., Goodman, G.L., Rajnak, K., Rana, R.S.: A systematic analysis of the spectra of the lanthanides doped into single crystal LaF_3 . *J. Chem. Phys.* **90**, 3443 (1989)
2. Yen, W.M., Hajime, Y. (eds.): *Phosphor Handbook*. CRC Press, Boca Raton (2012)
3. Xie, R.J., Li, Y.Q., Hirotsaki, N., Yamamoto, H.: *Nitride Phosphors and Solid-State Lighting*. Taylor & Francis, Boca Raton (2011)

Photoalignment

- [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

Photodetector

Bao-Jen Pong and Chao-Hua Wen
Center for Measurement Standards, Industrial
Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Light sensors](#); [Photosensors](#)

Definition

Photodetectors or photosensors are sensors that are used to transform incoming optical energy into electrical signals. A photodetector used in the application field of color science and technology typically covers the visible spectrum wavelength range from 380 to 780 nm.

Overview

Photodetectors are transducers that transform one of their characteristics when light energy impinges on them. In this category, photoresistors alter their ohmic resistance, rods, and cones of the retina neurons of the eye alter their electrochemical response, and chlorophyll in plant leaves changes the rate of converting CO_2 to O_2 . Some other photodetectors alter the flow of electrical current or the potential difference across their terminals.

Photodetectors with sufficiently fast response that provide a measurable output for a small amount of light are easily reproducible and are economical and worth investigating for applications in high-speed optical communications. This category includes avalanche photodiodes (APDs) and positive intrinsic negative photodiodes (PINs).

Specifications of photodetectors are characterized by key parameters as follows, spectral response, photosensitivity, quantum efficiency, dark current, forward-biased noise, noise equivalent power, terminal capacitance, timing response, frequency bandwidth, and cutoff frequency.

Figure 1 shows a typical photodetector. It includes three stages: (1) to generate electron–hole pairs or electrons in photosensitive material using an incident light; (2) to transform them through a bias provided by a power supply; and (3) to output the electrical signal directly or via an amplifier [1–5].

Photodiode and Photomultiplier Tube

Figure 1 shows a photodiode, a circuit including a p–n junction semiconductor and can output electrical signal. It can be seen that the photosensitive

material includes a p–n junction, a boundary or interface between n-type and p-type semiconductors, and a depleted semiconductor region with a high electric field that can separate the electron–hole pair.

As incident light on a photocathode (photosensitive material is made of a multi-alkaline material, for example), it generates photoelectrons; next, the generated photoelectrons travel in a vacuum tube and hit the first dynode and generate secondary electrons. Then, the secondary electrons are multiplied by a series of

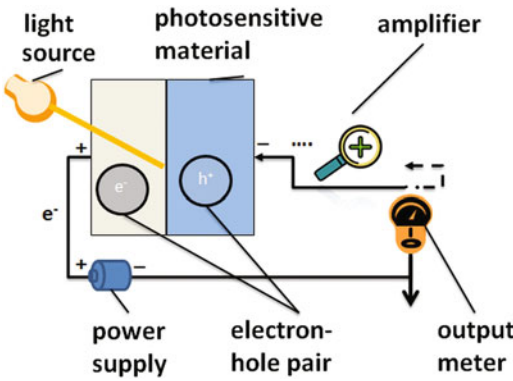
dynodes and reach an anode to produce an electrical signal. As the electrical signal is multiplied many times and the whole process is performed in a vacuum tube to prevent electron absorption by the air, it is called a photomultiplier tube (PMT), as shown in Fig. 2.

Photodiode Array and Charge-Coupled Device

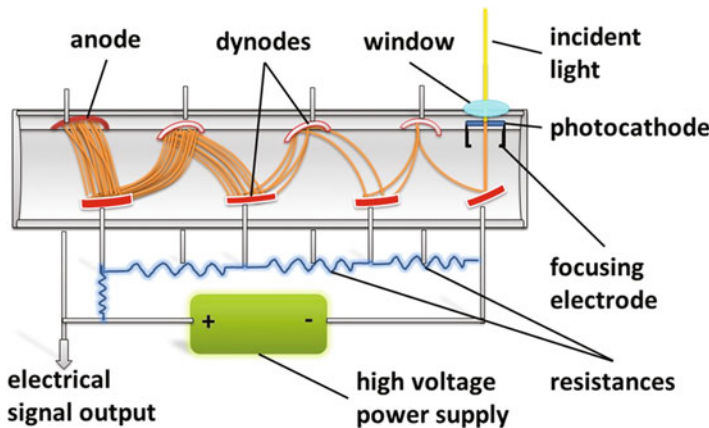
As the discrete photodiodes form an array with an integrated circuit board, it is called a photodiode array. As an integrated circuit contains an array of capacitors that store charge of electron–hole pairs created by photosensitive material, it is called a charge-coupled device (CCD).

Applications

In color applications, photodetectors are included in many color measuring instruments, which measure parameters such as luminance, illuminance, or chromaticity. Sze and Ng divided photodetectors into four categories: photoconductor, photodiode, avalanche photodiode, and phototransistor [1]. The photodiode and the phototransistor are typically used in color applications.

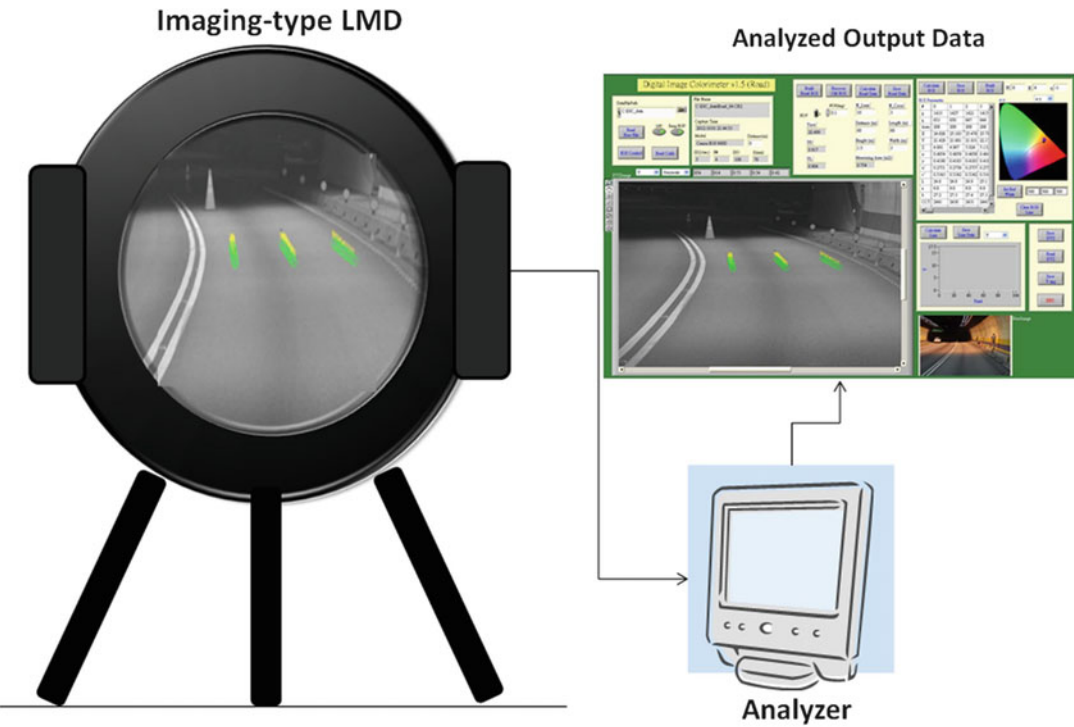


Photodetector, Fig. 1 A schematic diagram of a photodetector (i.e., also a photodiode)



Photodetector, Fig. 2 A schematic diagram to show a photomultiplier tube. The incident light through the window hits a photocathode (photosensitive material is made by a multi-alkaline material, for example). It generates photoelectrons, which transmit in a vacuum, hit on the first dynode, and is then multiplied by a series of dynodes,

finally reaching an anode to produce an amplified electrical signal. To prevent the absorption of photoelectrons and generated electrons into the air, a vacuum tube is used. Finally, the output electrical signal is amplified 10^4 – 10^5 times and within in a pico-ampere to micro-ampere current range



Photodetector, Fig. 3 The schematic diagram of an image luminance measuring device (ILMD). The captured color image via an ILMD, and then transformed into

a luminance image through the analyzer, and the luminance distribution can be used for many applications, such as to calculate the glare index for lighting applications

An Si photodiode together with a filter to mimic the luminous efficiency function (i.e., $V(\lambda)$ filter) becomes a $V(\lambda)$ photodetector, which is used for measuring the luminance (in cd/m^2 units). In a colorimeter, an Si photodiode combined with filters to mimic the color-matching functions of $\bar{x}$, $\bar{y}$, and $\bar{z}$ are used to measure CIE tristimulus values (X, Y, and Z).

A PMT can detect dim light; it usually works with a monochromator. This arrangement can measure spectral information across the spectrum. The measured spectra multiplying the luminous efficiency function $V(\lambda)$ can obtain luminance.

In color applications, a photodiode array is placed in an array image plane to form an image-based colorimeter. A CCD combined with the desired color filters measures an array of chromaticity values. Recently, in the display and imaging industry, CCD has been replaced by a complementary metal-oxide semiconductor (CMOS) as

an image-type photodetector because of its low cost. Figure 3 shows an image luminance measurement device (ILMD) that applies a commercial color CCD camera or color CMOS camera to measure the quality of roadway lighting [6]. ILMD is widely used for measuring luminaires of tunnel lighting, outdoor day-lite, indoor lighting (uniformity, contrast, glare, etc.), and near-field goniophotometry as well.

Cross-References

- [Goniophotometer](#)
- [Instrument: Colorimeter](#)
- [Instrument: Spectrophotometer](#)
- [Luminous Efficiency Function](#)
- [Monochromator](#)
- [Spectroradiometer](#)

References

1. Sze, S.M., Ng, K.K.: *Physics of Semiconductor Devices*, 3rd edn. Wiley, Hoboken (2007)
2. Razeghi, M.: *Fundamentals of Solid State Engineering*, 4th edn. Springer, New York (2019)
3. Omnes, F.: Introduction to semiconductor photodetectors. In: Decoster, D., Harari, J. (eds.) *Optoelectronic Sensors*. ISTE, London (2009)
4. Luna-López, J.A., Aceves-Mijares, M., Carrillo-López, J., Morales-Sánchez, A., Flores-Gracia, F., Vázquez Valerdi, D.E.: UV-vis photodetector with silicon nanoparticles. In: Gateva, S. (ed.) *Photodetectors*. InTechOpen, London (2012)
5. Ready, J.: Optical detectors and human vision. In: Roychoudhuri, C. (ed.) *Fundamentals of Photonics*, pp. 211–248. SPIE Press, Bellingham (2008)
6. Hsu, S.W., Chung, T.Y., Pong, B.J., Chen, Y.C., Hsieh, H.P., Lin, M.W.: Relations between flicker, glare, and perceptual ratings of LED billboards under various conditions. In: *Proceeding of CIE Century Conference “Towards a New Century of Light,” CIE x038:2013*, Paris, pp. 428–434 (2013)

Photonics

► [Ferroelectric Liquid Crystal Cells for Modern Applications in Displays and Photonics](#)

Photoreceptor Cells

► [Photoreceptors, Color Vision](#)

Photoreceptors, Color Vision

Marisa Rodriguez-Carmona¹ and
Emily J. Patterson²

¹Optometry and Visual Science, City University,
London, London, UK

²Institute of Ophthalmology, University College
London, London, UK

Synonyms

[Cone pigments](#); [Opsin genes](#); [Photoreceptor cells](#);
[Visual pigments](#)

Definition

The expression and molecular structure of the genes encoding the human cone photopigments is the first step in determining our color vision. Alterations in these genes have consequences for the spectral sensitivity of the cone photoreceptors, which in turn affect how color-normal and color-deficient individuals see the world.

Overview

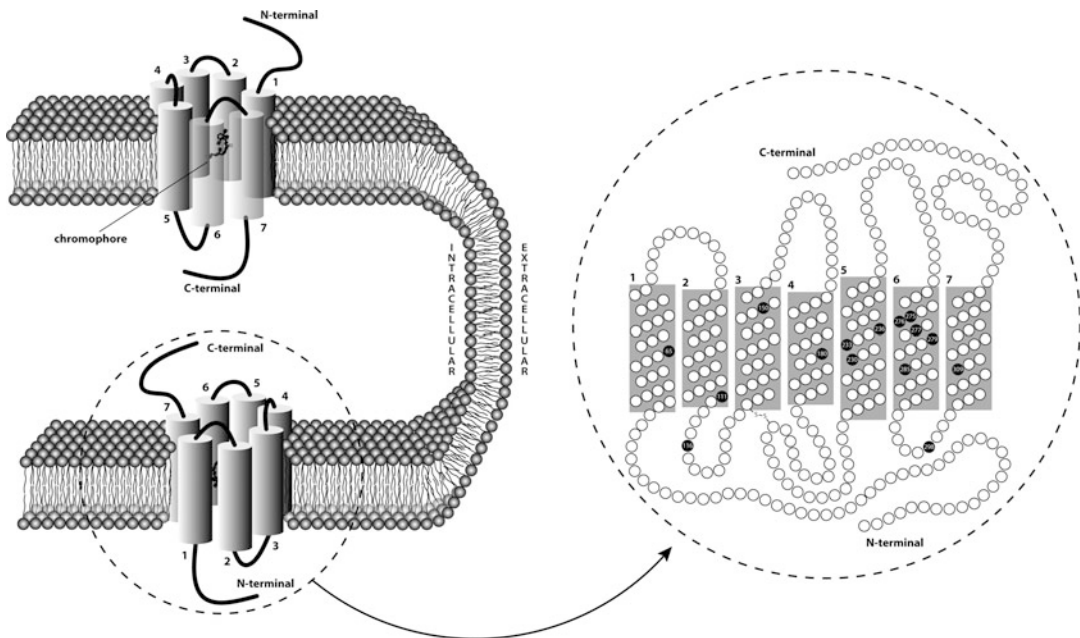
Normal trichromats can match any given color by combining three suitably chosen primaries (such as red, green, and blue) in appropriate proportions. Color-deficient individuals differ from normal trichromats with respect to the proportions of the primaries required for a match: they may use the three primaries in unusual quantities or they may only use two primaries. The first type of variation is known as anomalous trichromacy, which is assumed to arise when three classes of cones are present, one of which contains a photopigment with an anomalous absorption spectrum. The second type of variation is known as dichromacy, which occurs if only two of the three classes of cone – short (S), middle (M), or long (L) wavelength-sensitive – may be affected, manifesting in different color vision phenotypes. The different phenotypes are associated with predictable patterns of color-matching behavior, which can be identified easily using psychophysical tests of color vision. The presence and sequence of the genes encoding the photopigment genes ultimately determine the individual's genotype and can be identified by examining their deoxyribonucleic acid (DNA). Research into the empirical relation between underlying genotype and expressed phenotype has led to improved understanding of the physiological mechanisms responsible for variations in human color vision.

Nathans and colleagues [1] were the first to utilize advances in molecular genetics to study human photopigment genes, using a technique that involves in vitro construction of DNA molecules to produce recombinant DNA. Photopigments are the light-absorbing

molecules that mediate vision, with absorption maxima (λ_{max}) at approximately 420 nm (S-cone), 530 nm (M-cone), and 560 nm (L-cone). Structurally, they consist of opsin – transmembrane heptahelical proteins of a single polypeptide chain, composed of either 364 (*OPN1LW* and *OPN1MW* for the L- and M-cones, respectively) or 348 (*OPN1SW* for the S-cone) amino acids – bound to a chromophore: 11-*cis* retinal. Photon absorption by the pigment molecules causes isomerization of the chromophore from “11-*cis*” to “all-*trans*” form, which initiates a signal cascade, known as photo-transduction. The binding socket site of the chromophore in both the cone and rod opsins is located in helix 7 (Fig. 1).

The precise λ_{max} of a given pigment will be determined by the amino acid sequence of the opsin and the interaction of specific amino acids with retinal [3]. Whereas the λ_{max} of 11-*cis*

retinal is 375 nm [4, 5], the exact amino acid sequence of opsin “spectrally tunes” the pigment to a specific wavelength [6, 7] by shifting it toward the “red” end of the spectrum. Research with retinal organoids suggests that the fate of a given cone cell is determined by thyroid hormone signaling, with specification of S-cones occurring early in development and the remaining cones being specified as L vs M later [8]. The number and distribution of the three cone classes are not equal, either between individuals or across a given retina. In the normal eye, S-cones comprise approximately 7–10% of the total cone population and are even sparser at the fovea [9, 10] which may account for some of the variation in chromatic sensitivity observed across the retina [11]. The remaining cones are specified as L or M, the proportion of which, unlike S-cones, is highly variable and has been shown to range from around 50–95% L-cones in the normal eye [12].

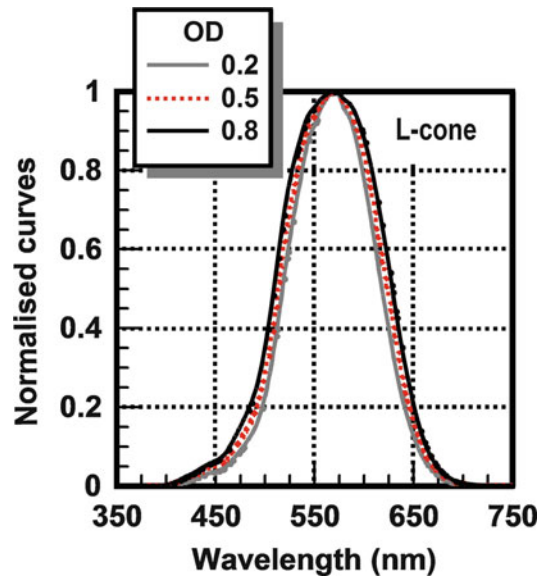


Photoreceptors, Color Vision, Fig. 1 Arrangement of visual pigments in a cone photoreceptor. The infolding membrane of the cone outer segment is packed with photopigment molecules. The molecule consists of seven α -helices which span the membrane of the cell surrounding the chromophore: 11-*cis* retinal. The NH₂ (N) terminus is extracellular, whereas the COOH (C) terminus is intracellular. Shown in the bottom right is the sequence of amino

acids that make up this heptahelical molecule. White circles represent amino acids that are identical across the L- and M-cone pigments, and black circles represent amino acids that are different. The substitutions at positions 180, 277, and 285 are believed to contribute most of the spectral difference between the M- and L-cone pigments. (Adapted from Sharpe [2] and Nathans [1])

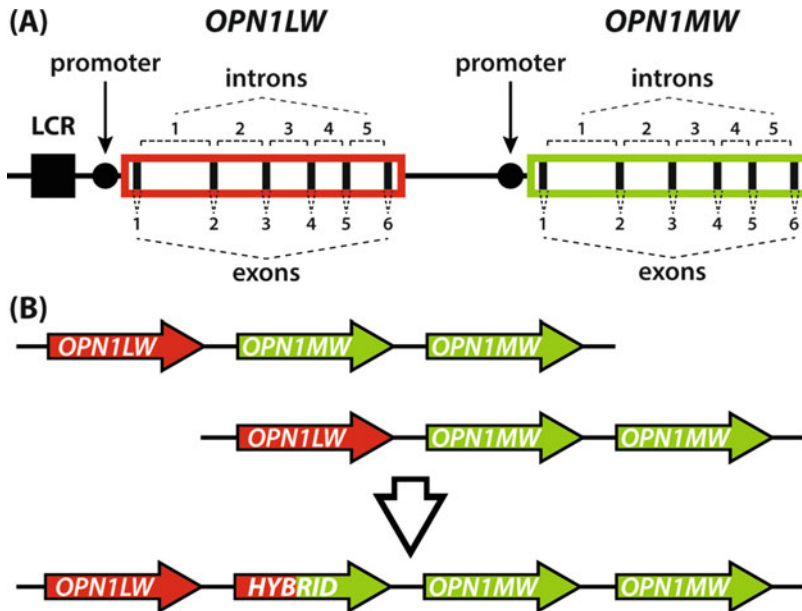
The genes encoding the S-cone pigments reside alone as a single copy on the q-arm (long arm) of chromosome 7 [13]. The genes encoding the L- and M-cone pigments reside on the q-arm of the X-chromosome, organized in a tandem array [1] of up to six exons (coding regions of DNA) separated by relatively long introns (non-coding regions of DNA) (Fig. 3). The S-cone pigment shows only $43 \pm 1\%$ analogy in the amino acid sequence when compared to either the M- or L-cone photopigments. In contrast, the M- and L-cone pigment genes are 96% homologous, differing only in 15 amino acids [2]. The approximate 30 nm spectral difference between these two visual pigments is attributable to substitutions at these particular amino acid positions confined to exons 2–5. The largest shifts in λ_{max} are produced by substitutions at two key sites within exon 5: amino acid positions 277 (~ 7 nm) and 285 (~ 14 nm) (Fig. 1). These result in spectral shifts of 15–25 nm; the exact value depends on sequences in exons 2–4. Substitutions within exons 2–4 produce much smaller shifts of less than ~ 4 nm and may be responsible for the subtle individual differences in color vision among normal and anomalous trichromats. Both in vitro methods to investigate protein sequence variation [1, 6] and in vivo methods to probe spectral sensitivity [14] have shown that amino acid substitutions in exon 2 contribute very little to spectral tuning (0.0–0.7 nm) [6, 14]. More recently, it has been speculated that the amino acid differences in exon 2 are involved in controlling the optical density of the L- and M-cone photopigments [15]. The optical density depends on the concentration of the photopigment in the cone outer segment, the length of the cone outer segment, and also the extinction coefficient, which describes the probability of a photon being absorbed [16]. An increase in effective optical density of cones causes a broadening of the spectral sensitivity curve (Fig. 2), such that if two genes that differ only in their exon 2 sequences are expressed, the resulting differences in the optical density of the photopigments may account for some residual color discrimination [15].

The high homology of the L- and M-cone pigment genes suggests evolution from a single



Photoreceptors, Color Vision, Fig. 2 Theoretical spectral sensitivity curves for the L-cone pigment. Changes in the optical density of the cone pigment may occur from amino acid substitutions in exon 2, altering the stability of the molecule or the efficiency with which it absorbs light. For wavelengths near the peak, the optical density qualitatively mimics the difference that would be produced by a spectral shift

gene. Comparison of amino acid sequences indicates that S-cones and rod receptors arose first from a common ancestral receptor. From comparisons with contemporary New World monkeys, which have only two cone photopigments, it is thought that a long wavelength gene duplicated and diverged to originate the L- and M-cone photopigments, at the time when Old World monkeys (trichromatic) separated from New World monkeys [1, 17]. The location of the *OPN1LW* and *OPN1MW* genes on the X-chromosome can account for the larger number of (red/green) color-deficient men compared to women, as men have only one X-chromosome. The head to tail arrangement of the *OPN1LW/OPN1MW* genes makes them susceptible to mispairing during meiosis, leading to unequal crossing over between gene arrays. If the crossover occurs between genes, this will result in the deletion of a gene from one chromosome and its addition to the other, whereas a crossover within a gene will lead to the production of a hybrid gene that



Photoreceptors, Color Vision, Fig. 3 Schematic of the tandem array of L- and M-cones on the q-arm of the X chromosome. (a) The LCR (locus control region) can activate only one of the promoter regions just upstream of a gene. The promoters are regulatory units upstream (or to the left) of the transcription site and regulate the rate of DNA transcription into RNA and hence the amount of

opsin gene expression. The opsin-coding sequences are divided into exons (black bars). The intron sequences (gaps) are silent or noncoding and have no apparent function. (b) Schematic of unequal intragenic crossover that would produce hybrid genes. (Adapted from Neitz and Neitz [18])

combines regions of the L and M genes into a single gene. Such hybrid genes are thought to be the genetic basis responsible for the majority of color vision defects [1].

Figure 3 shows an example of how the formation of hybrid genes produces pigments with abnormal spectral sensitivity. The spectral sensitivity of the photopigment will depend on which of the *OPN1LW*/*OPN1MW* genes the crossover originated from [19]. Visual pigment gene arrays for people with normal color vision have an *OPN1LW* gene in the most upstream (or leftmost) position, with one or more *OPN1MW* genes downstream (or to the right) of the *OPN1LW* gene. Normal trichromats usually have only one copy of *OPN1LW*, multiple copies of *OPN1MW*, and possibly a number of hybrids [2]. There is good evidence from studies of gene expression in the retina and from studies involving genotype-phenotype relationships [20] that

only the first two genes are expressed and play a significant role in color vision. Whereas normal trichromats have *OPN1LW* and *OPN1MW* occupying the two most proximal positions of the array from the locus control region (LCR), anomalous trichromats instead have a hybrid gene positioned in either the first or second position. Deutan deficiencies arise when there is no M-sensitive photopigment, whereas protan deficiencies arise from a lack of L-sensitive photopigment. If there is only one opsin gene, or two with identical spectral sensitivity, the resulting phenotype will be dichromacy (deuteranopia or protanopia). If, however, the first two genes in the array encode opsins with different spectral sensitivity, as can be introduced by a hybrid, the result will be anomalous trichromacy (deuteranomaly or protanomaly) [7]. The propensity for the *OPN1LW* and *OPN1MW* genes to cross over creates the opportunity for multiple combinations, producing

numerous polymorphisms and hybrid variants, which are thought to underlie the high variability in red/green color vision observed even within the normal population [18].

In addition to the creation of hybrid genes, mutations within the *OPN1LW/OPN1MW* array can produce cones that are nonfunctional and even degenerative, depending on the amount of correctly spliced precursor messenger ribonucleic acid (RNA) [21]. High-resolution retinal imaging has shown that those with genes encoding functional (even if altered) opsin in one or both proximal positions of the array present with a contiguous mosaic of cone photoreceptors [22]. As such, other non-chromatic aspects of visual function in single or multigene dichromats are usually unaffected [23]. In contrast, retinal imaging in those with genes encoding nonfunctional opsin reveals variable disruption of the cone mosaic [24], which may be associated with other visual deficits, such as myopia [25].

The *OPN1SW* opsin gene sequence is nearly invariant in the human population; however further genotype-phenotype studies are yet to be performed to prove this. Intragenic crossover – the mechanism that permits the frequent manifestation of anomaly in protan and deutan defects – has no analogy in tritan defects. Six mutations have been established causing amino acid substitutions that perturb the structure or stability of the S-cone photopigment [13, 26], therefore strongly affecting the performance of the S-cone photoreceptors. On the other hand, no polymorphisms causing shifts in $\lambda_{\max}$ have been reported so far, and only one substitution was found in the coding sequences and exon-intron junctions, which has no impact on its spectral characteristics [27]. However, abnormalities as a result of genetic variation in *OPN1SW* may have less likelihood of being detected, as any small variations in $\lambda_{\max}$ would be difficult to dissociate from individual variations in macular pigment and lens density measured psychophysically or in vivo. Testing patients with blue cone monochromacy, for whom cone function is isolated to the S-cones [28], may enable detection of individual differences in S-cone spectral sensitivity, although its

low prevalence (~1/100,000) limits the opportunities for such research studies.

Future Directions

An important consideration in the study of human color vision is how spectral differences between photopigments translate to performance on color-related tasks. There are numerous retinal and cortical factors that influence chromatic discrimination; it is therefore insufficient to simply compare differences in $\lambda_{\max}$. For example, there have been shown to be large differences in L:M cone ratio [12] even among family members with the same genotype [25]; how such differences could interact with hybrid genes and mutations to affect color vision is not well understood. When color vision depends on subtle differences between L- and M-cone pigments, as in anomalous trichromacy, polymorphisms that might affect optical density as well as those that shift $\lambda_{\max}$ must be accounted for. It has been suggested that a L/M separation greater than ~20 nm would be sufficient to account for normal color vision, implying that normal trichromats can, in principle, have visual pigments encoded by hybrid genes, providing that the $\lambda_{\max}$ values of the two hybrids remains greater than 20 nm apart.

Changes in $\lambda_{\max}$, photoreceptor optical densities, and/or post-receptoral amplification of cone signals appear to be the most important parameters that affect our red-green chromatic sensitivity. In addition to cone photoreceptors, rods and melanopsin also form part of the process [29], particularly under certain low lighting conditions [30] or in color deficiency or cone dysfunction [31, 32]. Establishing the relative importance of each of these parameters in determining an individual's color discrimination performance therefore remains a difficult task. A marriage of information derived from genetic analysis of pigment genes, in vitro and in vivo characterization of cone cells, psychophysical data on chromatic sensitivity, and theoretical modeling is therefore needed in order to produce a comprehensive

description of mechanisms that underlie human color vision in all its complexity.

Cross-References

- [Chromophore](#)
- [Dichromacy](#)
- [Trichromacy](#)

References

1. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B., Hogness, D.S.: Molecular genetics of inherited variation in human color vision. *Science*. **232**(4747), 203–210 (1986)
2. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Opsin genes, cone photopigments, color vision, and color blindness. In: Gegenfurtner, K.R., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 3–52. Cambridge University Press, Cambridge (1999)
3. Bowmaker, J.K.: Visual pigments and molecular genetics of color blindness. *News Physiol. Sci.* **13**, 63–69 (1998)
4. Knowles, A., Dartnall, H.J.A.: The photobiology of vision. In: Davson, H. (ed.) *The Eye*, vol. 2B, pp. 562–569. Academic, New York (1977)
5. Nakanishi, S.: Mammalian tachykinin receptors. *Annu. Rev. Neurosci.* **14**(1), 123–136 (1991)
6. Asenjo, A.B., Rim, J., Oprian, D.D.: Molecular determinants of human red/green color discrimination. *Neuron*. **12**(5), 1131–1138 (1994)
7. Neitz, M., Neitz, J., Jacobs, G.H.: Spectral tuning of pigments underlying red-green color vision. *Science*. **252**(5008), 971–974 (1991)
8. Eldred, K.C., Hadyniak, S.E., Hussey, K.A., Brennerman, B., Zhang, P.W., Chamling, X., et al.: Thyroid hormone signaling specifies cone subtypes in human retinal organoids. *Science*. **362**(6411), eaau6348 (2018)
9. Ahnelt, P.K., Kolb, H., Pflug, R.: Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *J. Comp. Neurol.* **255**(1), 18–34 (1987)
10. Curcio, C.A., Allen, K.A., Sloan, K.R., Lerea, C.L., Hurley, J.B., Klock, I.B., Milam, A.H.: Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* **312**(4), 610–624 (1991)
11. Williams, D.R., MacLeod, D.I., Hayhoe, M.M.: Punctate sensitivity of the blue-sensitive mechanism. *Vis. Res.* **21**(9), 1357–1375 (1981)
12. Hofer, H., Carroll, J., Neitz, J., Neitz, M., Williams, D. R.: Organization of the human trichromatic cone mosaic. *J. Neurosci.* **25**(42), 9669–9679 (2005)
13. Weitz, C.J., Went, L.N., Nathans, J.: Human tritanopia associated with a third amino acid substitution in the blue-sensitive visual pigment. *Am. J. Hum. Genet.* **51**(2), 444–446 (1992)
14. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., et al.: Red, green, and red-green hybrid pigments in the human retina: correlations between deduced protein sequences and psychophysically measured spectral sensitivities. *J. Neurosci.* **18**(23), 10053–10069 (1998)
15. Neitz, J., Neitz, M., He, J.C., Shevell, S.K.: Trichromatic color vision with only two spectrally distinct photopigments. *Nat. Neurosci.* **2**(10), 884–888 (1999)
16. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York (1982)
17. Nathans, J.: The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron*. **24**(2), 299–312 (1999)
18. Neitz, M., Neitz, J.: Numbers and ratios of visual pigment genes for normal red-green color vision. *Science*. **267**(5200), 1013–1016 (1995)
19. Neitz, J., Neitz, M., Kainz, P.M.: Visual pigment gene structure and the severity of color vision defects. *Science*. **274**(5288), 801–804 (1996)
20. Hayashi, T., Motulsky, A.G., Deeb, S.S.: Position of a ‘green-red’ hybrid gene in the visual pigment array determines colour-vision phenotype. *Nat. Genet.* **22**(1), 90–93 (1999)
21. Greenwald, S.H., Kuchenbecker, J.A., Rowland, J.S., Neitz, J., Neitz, M.: Role of a dual splicing and amino acid code in myopia, cone dysfunction and cone dystrophy associated with L/M opsin interchange mutations. *Transl. Vis. Sci. Technol.* **6**(3), 2 (2017)
22. Wagner-Schuman, M., Neitz, J., Rha, J., Williams, D. R., Neitz, M., Carroll, J.: Color-deficient cone mosaics associated with Xq28 opsin mutations: a stop codon versus gene deletions. *Vis. Res.* **50**(23), 2396–2402 (2010)
23. Jägle, H., de Luca, E., Serey, L., Bach, M., Sharpe, L.T.: Visual acuity and X-linked color blindness. *Graefes Arch. Clin. Exp. Ophthalmol.* **244**(4), 447–453 (2006)
24. Carroll, J., Rossi, E.A., Porter, J., Neitz, J., Roorda, A., Williams, D.R., Neitz, M.: Deletion of the X-linked opsin gene array locus control region (LCR) results in disruption of the cone mosaic. *Vis. Res.* **50**(19), 1989–1999 (2010)
25. Patterson, E.J., Wilk, M., Langlo, C.S., Kasilian, M., Ring, M., Hufnagel, R.B., et al.: Cone photoreceptor structure in patients with X-linked cone dysfunction and red-green color vision deficiency. *Invest. Ophthalmol. Vis. Sci.* **57**(8), 3853–3863 (2016)
26. Baraas, R.C., Hagen, L.A., Dees, E.W., Neitz, M.: Substitution of isoleucine for threonine at position

- 190 of S-opsin causes S-cone-function abnormalities. *Vis. Res.* **73**, 1–9 (2012)
27. Crognale, M.A., Teller, D.Y., Yamaguchi, T., Motulsky, A.G., Deeb, S.S.: Analysis of red/green color discrimination in subjects with a single X-linked photopigment gene. *Vis. Res.* **39**(4), 707–719 (1999)
 28. Neitz, J., Neitz, M.: The genetics of normal and defective color vision. *Vis. Res.* **51**(7), 633–651 (2011)
 29. Lucas, R.J., Lall, G.S., Allen, A.E., Brown, T.M.: How rod, cone, and melanopsin photoreceptors come together to enlighten the mammalian circadian clock. *Prog. Brain Res.* **199**, 1–18 (2012)
 30. Zele, A.J., Maynard, M.L., Joyce, D.S., Cao, D.: Effect of rod–cone interactions on mesopic visual performance mediated by chromatic and luminance pathways. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **31**(4), A7–A14 (2014)
 31. Kurtenbach, A., Meierkord, S., Kremers, J.: Spectral sensitivities in dichromats and trichromats at mesopic retinal illuminances. *J. Opt. Soc. Am. A.* **16**(7), 1541–1548 (1999)
 32. Luo, X., Cideciyan, A.V., Iannaccone, A., Roman, A. J., Ditta, L.C., Jennings, B.J., et al.: Blue cone monochromacy: visual function and efficacy outcome measures for clinical trials. *PLoS One.* **10**(4), e0125700 (2015)

Photosensitive Retinal Ganglion Cells

- [Melanopsin Retinal Ganglion Cells](#)

Photosensors

- [Photodetector](#)

Physiologic Scotoma

- [Blind Spot](#)

Piezoelectric Ink-jet (PIJ)

- [Coloration, Ink-Jet Printing](#)

Pigment, Ceramic

Guillermo Monros

Departamento de Química Inorgánica y Orgánica,
Universitat Jaume I, Castelló, Spain

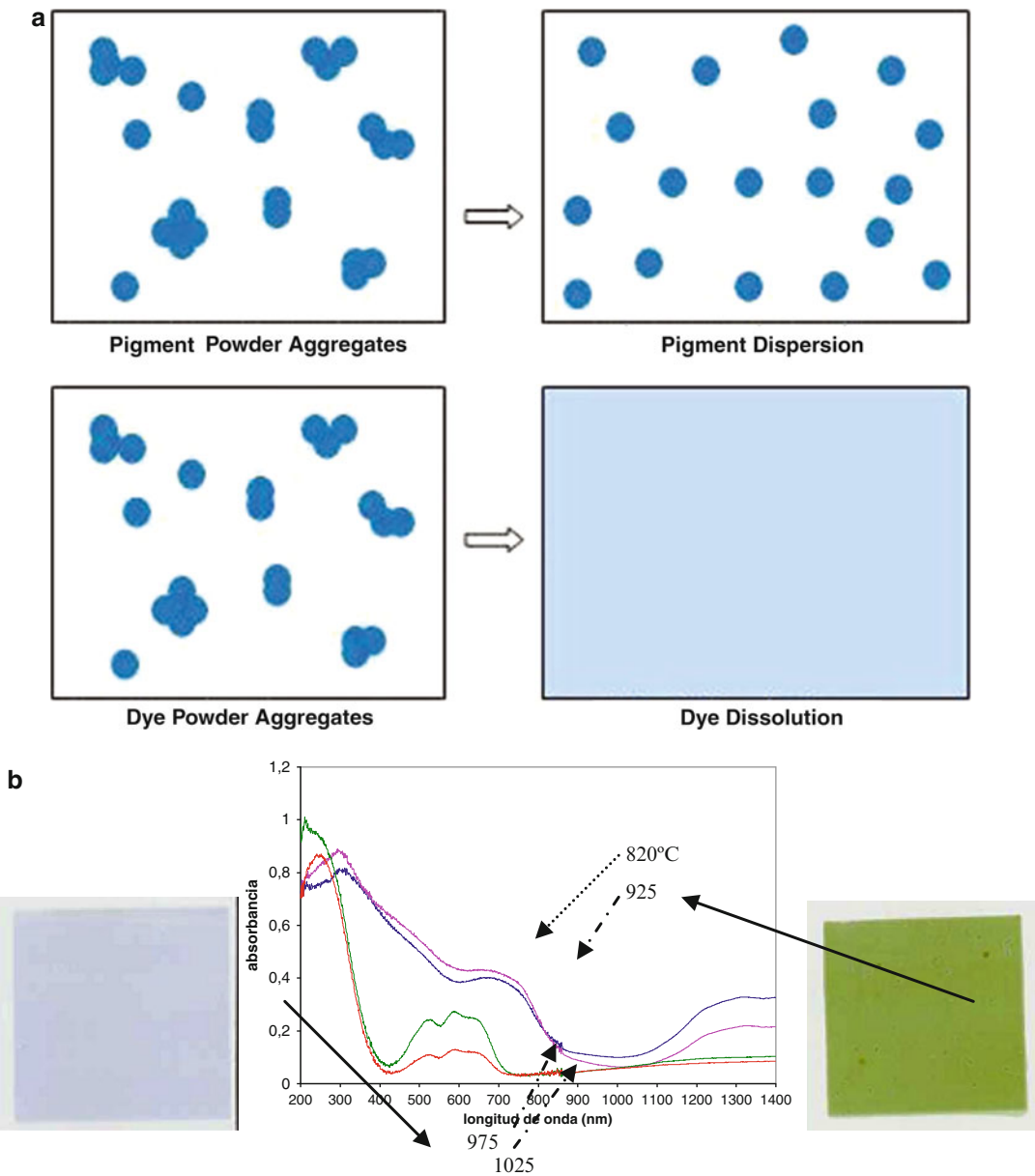
Synonyms

[Colorant](#); [Dye](#); [Paint](#); [Stain](#); [Tint](#)

Definition

A ceramic pigment is usually a metal transition complex oxide obtained by a calcination process which shows three main characteristics: (a) thermal stability, maintaining its identity when temperature increases; (b) chemical stability, maintaining its identity when fired with glazes or ceramic matrices; and (c) high tinting strength when dispersed and fired with glazes or ceramic matrices. Other characteristics as high dispersability in vehicles, high refractive index (in order to avoid transparency and to increase its tinting strength), acid and alkali resistance, and low abrasive strength are also suitable.

The main coloring methods of ceramics are based on dyes or pigments. Strictly a dye or soluble colorant is a colored substance that interacts with the matrix to which it is applied and usually is soluble in the media of application called matrix or substrate. Conversely, ideally a pigment or stain is a colored substance insoluble in the pigmented matrix, is stable in the matrix (it does not interact with the matrix), and, in the case of ceramic pigments used in stoneware or glazes, is stable when temperature increases (thermal stability). Obviously a pigment must have a high tinting strength relative to the materials it colors. As a general rule, ceramic stains and ceramic pigments look pretty much the same before and after firing, but not dyes, based on raw oxides as well as salts such as carbonates or nitrates, because they decompose in firing and dissolve in the glazes or stoneware (Fig. 1a) [1a].



Pigment, Ceramic, Fig. 1 (a) Ideal behavior of interaction of pigments (dispersion) and dyes (dissolution) with the matrix. (b) Color evolution with the firing temperature of screen printing (90 threads/cm) deposition of a Co-Ti

ink (1 w.% in the organic polyol media) under a glazed CaO-ZnO-SiO₂ monoporosa tile (previously fired at 1,080 °C)

Industrial Ceramic Pigments

In the industry there are several steps for the production of industrial ceramic pigments:

- 1. Grinding. The raw materials (usually oxides or carbonates) are carefully mixed in mills such as ball mills often with mineralizer agents (inorganic salts as NaCl, NaF, and KNO₃) in order to facilitate the pigment crystallization.

2. Calcination. The mixture is then calcined in either batch kilns or continuous calciners.
3. Washing. After calcination, the powders are washed to eliminate residual salts from mineralizers and other leaching components in order to avoid future firing defects and environmental affections.
4. Micronization. The mixture is then ground to the necessary fineness in mills. Micronizers and jet mills are used to break agglomerates.
5. Tinting strength control. The final production step involves careful control of the color tone.

Because these pigments are formed at high temperatures, they generally offer superb thermal stability and are relatively inert. This results in excellent weathering and light fastness properties. Most of these pigments have superior acid and alkali resistance. They are nonmigrating and nonbleeding in nature and do not interact with substrates. The principal disadvantage of ceramic pigments is their low tinting strength. In addition, some are relatively high in cost. This is particularly true with cobalt-containing pigments. Some of these pigments are difficult to disperse. However, the recent development of easily dispersed ceramic pigments should eliminate this problem. A final concern is the inherent hardness of these pigments. Their hardness can lead to processing system damage through abrasion. When using ceramic pigments, processing system components designed for use with abrasive materials should be considered [1b].

Overview

Stable fired pigments solve some of the problems found in using dyes such as: (a) the use of hazardous raw materials, because many of the unstable oxides or salts used as direct dyes are soluble or toxic, its calcination, by combining these elements along with clays, silica, alumina, etc., makes them stable giving a safety manipulation, is the case of vanadium pentoxide a toxic substance that is safely used in the zirconium vanadium blue zircon ceramic pigment, (b) volatilization in the kiln which is colored by

the fumes such as it occurs using vanadium oxide or chromium oxide, and (c) the own stability of the color avoiding chemical reactions of the dye with the components of the glaze, for example, eskolaite Cr_2O_3 , used as natural green dye, reacts with tin oxide and silica of the glazes producing malayaite Cr-CaSnSiO_5 pink precipitates, and then a pink undesired coloration is given, or reacts with zinc oxide of the glaze to crystallize ZnCr_2O_4 spinel, imparting undesired blue-brown color; the use of the cobalt-zinc-alumina-chromite blue-green pigment gives a trustworthy stable range of color from green to blue.

How to give trustworthy and safety color to ceramics is the main proposal of pigment production because adding color to our ceramics is needed to produce ceramic art and it cannot be a tricky proposition. Working with paints, the color you put on your piece is the color of the final paint, but in ceramics the fire also builds the color. Pigment research is addressed to obtain new ceramic pigments in order to reach three main purposes:

- (a) High thermal stability pigments, addressed to new materials as porcelainized stoneware and correlated use of ink-jet printing techniques, that require obtaining adequate particle size of pigments or precursor solutions that produce the pigment “in situ,” during thermal processing of substrate.
- (b) High chemical resistance pigments, adapted to new glazes and ceramic matrices avoiding particle pigment solubilization or degradation when it reacts in ceramic processing.
- (c) Eco-friendly pigments, with low environmental affection evaluated on its complete life cycle: raw material preparation, transport, processing, product distribution, and waste management.
- (d) Pigments for industrial digital printing (ink-jet). Ink-jet decoration is a relatively recent methodology. Elmquist designed the first practice application of CIJ (continuous ink-jet) in 1951 for tape recorder. Winston developed the first teletype printer by CIJ, and in 1968 the first commercial printer 9600 from AB Dick appeared. A continuous

flow of electrical conductive ink drops is thrown to substrate and selectively dispersed by an electric field in the CIJ ancient technology. The DOD (drop on demand) ink-jet was developed later in order to produce only the demanded drops. Firstly, in TIJ (thermal ink-jet) technique, the drops were produced from pressure of bubbles generated into the nozzles by a local heating mechanism. Now, in the PIJ (piezoelectric ink-jet) technique, the drops are produced by pressure due to the deformation when an electric field is applied to a piezoelectric piece disposed in the nozzles.

Digital decoration technology is widely developed in other materials (textile, paper, etc.), basically using TIJ technique. It has been claimed for surface ceramic tile decoration due to several advantages based on the fact of absence of contact between the applicator and the decorated surface:

- (i) High resolution of images. The ink-jet technique allows to throw 2,000–5,000 drops/cm in a strictly controlled way. Using serigraphy, chalcography, or flexography methods, the resolution is very limited. Therefore photographic quality images could be carefully reproduced by ink-jet digital monitored application.
- (ii) Low raw strength of tile and short enameling line required. Due to the absence of contact applicator surface, the raw strength is limited only to the adequate manipulation of raw tile, and then the thickness of piece can be reduced. The number of decoration operations is reduced sometimes at one because it can be applied four colors in a selective way at the same time; likewise the reduced weight of ink deposited and relatively low water concentration reduce the drying time between successive applications, and therefore the length of industrial enameling line is drastically reduced.
- (iii) The topography of the tile surface is not limited and relief surfaces can be easily decorated.

- (iv) Reduction of necessary ink (1 g/m^2), fully recycling and washing operations eliminated.
- (v) Simplicity of the decoration process because it can be limited to only one decoration step using the simultaneous four-color application CMYK (cyan, magenta, yellow, black).

The main difficulty for ink-jet application in ceramic tile industry is the ink that must show specific properties, basically:

- (i) High stability, because the precipitation, agglomeration, or viscosity changes can obstruct the nozzles.
- (ii) High color strength is necessary due to reduced weight of ink deposited by dropping; on contrary, diffused and weak colors are obtained and poor decoration is produced.
- (iii) Neutral pH is required in order to prevent corrosive effects on the nozzles, but conductivity is not required in both PIJ or TIJ technologies; contrarily in ancient CIJ technique, the electric charge capacity of drops was necessary for their selective dispersion by an electric field.

There are three main families of digital inks: (a) soluble salts such as cobalt nitrate (*CYAN*), mixture of chromium nitrate and iron nitrate (*MAGENTA*), mixture of antimony nitrate, nicke and stabilized titania (*YELLOW*), and mixtures of cobalt and iron nitrates (*BLACK*) are used for producing “in situ,” when decorated ceramics are fired, classical ceramic pigments such as Co_2SiO_4 olivine blue pigment, $\text{Fe}(\text{Fe},\text{Cr})_2\text{O}_4$ iron chromite brown spinel, $(\text{Ti},\text{Ni},\text{Sb})\text{O}_2$ chromium-antimony yellow rutile, and CoFe_2O_4 black spinel, respectively; (b) colloidal or ultra-micronized classical pigments particles dispersed using dispersant agents; and (c) nanoparticles of classical ceramic pigments, produced by “soft chemistry” such as polyol or sol-gel routes, or metals absorbed on stable sol particles of oxohydroxy compounds of Al, Ti, Sn or Zr (purple of Cassius and its analogous with Cu).

Since the nineteenth century, needs on color stability, glaze resistance, and sure management lead to stabilized fired pigments. On these pigmenting systems, chromophore agents (usually first row transition heavy metals and p metals) are inertized in high-thermal and low-solubility ceramic matrices.

The main mechanisms of origin of color based on the stabilization-inertization of chromophores are the following:

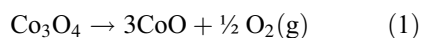
- (a) Structural mechanism. In this case the chromophore cations (usually transition metals that absorb, as it is discussed below, selected visible wavelengths by d-d electron transitions) are forming the ceramic network such as Co(II) in Co_2SiO_4 olivine. In other structural cases, a solid solution is obtained: pigmenting cations substitute structural cations in ceramic matrix as V^{4+} substitutes Zr^{4+} in ZrSiO_4 zirconium vanadium blue zircon (a cyan pigment also named Turkish blue zircon or turquoise zircon) formulated by CPMA as $(\text{V,Zr})\text{SiO}_4$. The above-discussed classical pigments $\text{Fe}(\text{Fe,Cr})_2\text{O}_4$ (iron chromite brown spinel with iron in solid solution into chromite network) and $(\text{Ti,Ni,Sb})\text{O}_2$ (chromium-antimony yellow rutile with Sb^{+5} and Cr^{3+} in solid solution into rutile lattice), are also solid solutions.
- (b) Inclusion mechanisms. Colored nanoparticles protected into high stable host crystals (e.g. Fe_2O_3 hematite into zircon crystal particles that gives the pink coral of iron in the zircon ceramic pigment) are the origin of color in this case.
- (c) Mordant mechanism. The origin of color is in this case a metal-based dye stabilized by a mordant compound. The final color depends on the mordant used, due to the formation of metal complexes that cause a change in the molecular orbital energies and hence a shift in UV-Vis absorption bands; e.g. Au in $\text{Sn}(\text{OH})_4$ mordant substrate on Cassius purple pigment [1a].

In agreement with above-described purposes of research on new ceramic pigments, three main

limitations can be pointed out: (a) thermal and chemical stability, (b) structure base, and (c) environmental safety. These limitations are described and illustrated with three examples.

Thermal and Chemical Stability: Cobalt Inks

Usually cobalt oxide Co_3O_4 used as cobalt precursor raw material is a mixed valence oxide of Co^{2+} and Co^{3+} with spinel structure $\text{Co}^{\text{II}}\text{Co}_2^{\text{III}}\text{O}_4$ which decompose on firing by reducing the unstable Co^{3+} to Co^{2+} producing a profuse pinhole in the glazes if the time of firing is not sufficient due to oxygen release:



The Co^{2+} ion dissolves in ceramic glazes and imparts an intense blue color to ceramic glazes up to 0.05 w.% and could be used in novel dye ink-jet application in ceramics. This color intensifies in the presence of ZnO, due to the crystallization of a solid solution of Co^{2+} in willemite $(\text{Zn}, \text{Co})_2(\text{SiO}_4)_2$, or in the presence of BaO which precipitates cobalt-celsian solid solution $(\text{Ba}, \text{Co})(\text{SiO}_4)_2$. In glazes containing MgO and TiO_2 , cobalt gives green colorations, due to crystallization of green cobalt-ilmenite CoTiO_3 more stable in boron glazes [1c].

The color evolution with the firing temperature of screen printing (90 threads/cm) deposition of a Co-Ti ink (1 w.% in the organic polyol media) over a glazed “monoporosa” tile (previously fired a 1,080 °C) is shown in Fig. 1b. The green coloration at low temperatures (820, 925 °C) is associated to cobalt-ilmenite CoTiO_3 crystallizations with Co^{2+} in octahedral coordination (absorption bands at 350 nm in the UV range, 700 nm in the visible red wavelength, and 1,200 nm in the IR range, in the spectral and color rendering of Fig. 1). When temperature increases, the cobalt-ilmenite CoTiO_3 decomposes and Co^{2+} solubilizes in the glaze giving intense blue color associated to Co^{2+} in tetrahedral coordination (absorption bands at 530 nm green, 590 nm orange, and 650 nm red and high reflectivity at

430 nm blue wavelength). At low temperature the color is associated to a pigment mechanism (particles of ilmenite crystallization on the glaze surface), but at high temperatures, the color is due to a dye mechanism (cobalt dissolution in glaze).

Structure Base: The Classical Pigment Palette

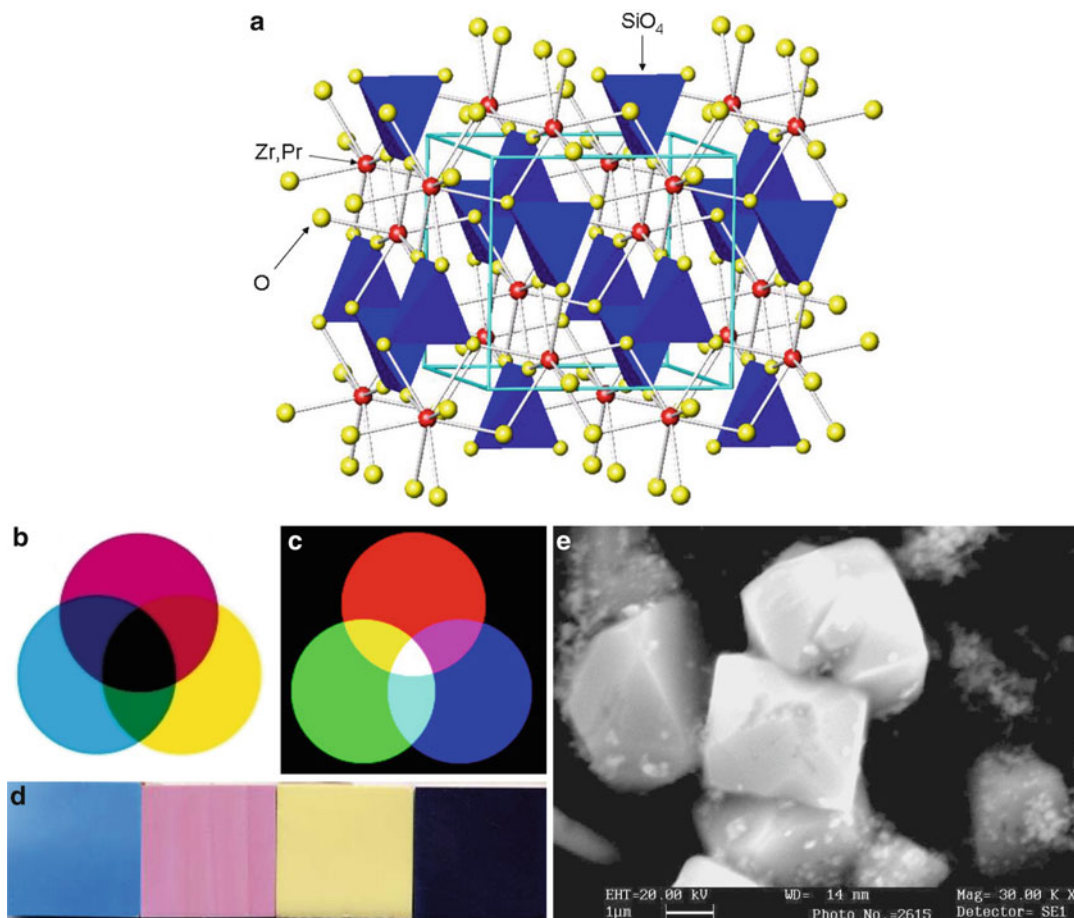
A useful colorant classification of the chemicals used as ceramic pigments was reached in 1977 from the requirements of the Toxic Substances Control Act, 94–469 US law, which includes the whole chemical substances used in the USA whether they are toxic or not. The Dry Colors Manufacturer's Association (DCMA) commends to their Metal Oxide and Ceramic Color Subcommittee of DCMA Ecology Committee the classification of the commercial ceramic pigments giving a standard terminology. This DCMA committee applied a chemical-structural criteria for the classification of ceramic pigments in 14 structural families: (1) baddeleyite [1 ceramic pigment], (2) borate [1], (3) corundum-hematite [4], (4) garnet [1], (5) olivine [2], (6) periclase [1], (7) phenacite-willemitite [1], (8) phosphate [2], (9) priderite [1], (10) pyrochlore [1], (11) rutile-cassiterite [11], (12) sphene [1], (13) spinel [19], and (14) zircon [3 ceramic pigments].

In 2010 the Color Pigments Manufacturers Association (CPMA) actualized the classification and chemical description of the complex inorganic color pigments in the fourth edition [2]. CPMA describes the industrial ceramic pigment by a code such as CPMA 1-01-4. The first number on the code is the corresponding number of the structure above discussed (in this case 1 = baddeleyite), the second set is the CPMA category number identifying each pigment within a given crystal class (in this case 1 = first ceramic pigment on the CPMA list), and the third set of numbers identifies the color of the pigment as follows: (1) violet and red-blue; (2) blue and blue-green; (3) green; (4) yellow and primrose; (5) pink, orchid, coral, and peach; (6) buff; (7) brown; (8) gray; and (9) black (in the above case 4 = yellow). Finally, CPMA also gives a

classification based upon the predominant use of these pigments; three use categories were adopted for guidance. Category A deals with pigments suspended in glass matrices which require the highest degree of heat stability and chemical resistance to withstand the attack of molten glass. Category B deals with pigments suspended in plastics and other polymers which require only moderate heat stability. Category C deals with pigments suspended in liquid vehicles which require little or no heat stability.

Among the more successful ceramic pigments which gives a complete CMYK palette of color, the pigments based on zircon structure (number XIV in CPMA classification) shown in Fig. 2a and discovered by C. A. Seabright in 1948 stand out.

- (a) CYAN based on the vanadium blue zircon $V\text{-ZrSiO}_4$ (CPMA number 14-42-2). CPMA describes this pigment as an inorganic pigment and as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and vanadium (IV) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(Zr,V)SiO_4$. Its composition may include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring ceramic glazes and clay bodies. Not generally used in porcelain enamels.
- (b) MAGENTA based on the pink coral $Fe\text{-ZrSiO}_4$ (CPMA number 14-44-5). CPMA describes the zirconium iron pink (peach, coral) zircon, as an inorganic pigment, as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and iron (III) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(Zr,Fe)SiO_4$. Its composition may include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring ceramic glazes. Not generally used in clay bodies or porcelain enamels.



Pigment, Ceramic, Fig. 2 (a) ZrSiO_4 structure from CPMA classification, S.G. $I4_1/$ with D_{2d} symmetry for SiO_4 group and dodecahedral D_{2d} for ZrO_8 , (b) CMYK subtractive color model, (c) RGB additive color model, (d) the CMYK zircon palette (*cyan* (Zr,V) SiO_4 DCMA

14-42-2, *pink* (Zr,Fe) SiO_4 DCMA 14-44-5, *yellow* (Zr,Pr) SiO_4 DCMA 14-43-4) including a Cr,Co-ferrite as *black* component, all 3% enameled in monoporosa glaze, (e) particles of black spinel pigment by SEM

P

(c) **YELLOW** based on the yellow of praseodymium Pr-ZrSiO_4 (CPMA number 14-43-4). CPMA describes the zirconium praseodymium yellow zircon, as an inorganic pigment, as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and praseodymium (III, IV) oxide (Pr_6O_{11}) in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(\text{Zr,Pr})\text{SiO}_4$. Its composition may include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring

ceramic glazes and clay bodies. Not generally used in porcelain enamels.

(d) **BLACK** usually obtained from spinel structure in CPMA classification such as the ferrite $(\text{Mg}_{0.5}\text{Fe}_{0.5})(\text{Co}_{0.5}\text{Fe}_{1.4}\text{Cr}_{0.1})\text{O}_4$ shown in Fig. 2d. CPMA describes the iron cobalt chromite black spinel as an inorganic pigment and as a reaction product of high-temperature calcination in which iron (II) oxide, cobalt (II) oxide, iron (III) oxide, and chromium (III) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of spinel. Basic chemical formula: $(\text{Co,Fe})(\text{Fe,Cr})_2\text{O}_4$. Its composition

may include any one or a combination of the modifiers Al_2O_3 , B_2O_3 , CuO , MnO , NiO , or SiO_2 . Use categories A, B, and C. Predominantly used for coloring category A substrates.

Adequate mixtures of above-described pigments may produce all range of visible colors, using the subtractive color vision model CMYK (where C = cyan, M = magenta, Y = yellow, and K = key associated to a mixture of C+M+Y which gives the black color) shown in Fig. 2b. in opposition to the additive color vision model RGB (R = red, G = green, B = blue) shown in Fig. 2c. Really the black color is not necessary because it could be built mixing CMY, but in ink and paints, it is used in order to simplify and to cheapen the printing manufacture. A detailed study of the CMY zircon pigments can be found in several doctoral theses [4, 5]. Really the color given by CMY zircon pigments is influenced by temperature and glaze or ceramic matrix; therefore usually a wider color palette must be used on pigmenting ceramics.

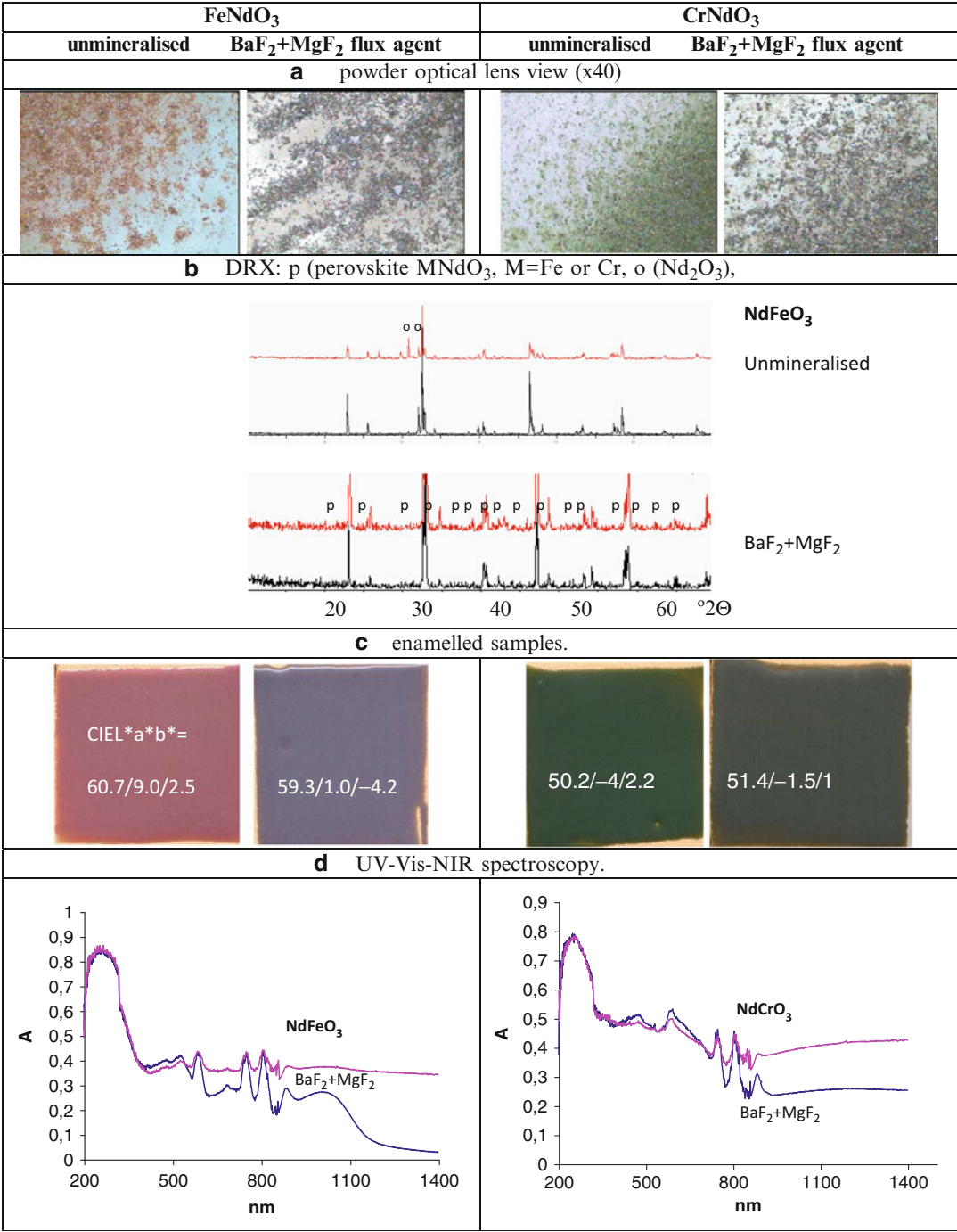
A total of 44 chemical substances are listed in the CPMA classification referred exclusively to industrial ceramic pigments. But the world of pigments is wider and, for example, on the industrial classification, the main red colors do not appear: the unstable purple of Cassius discovered by Johann Rudolph Glauber in 1659 and the sulfoselenide dated on 1909 and classified as harmful substance due to cadmium presence. The reddish ceramic pigments have been during centuries the driving force in the ceramic pigment field research [6]. The investigation on new host structures for pigmenting also has received attention, but usually, searching the red shades.

As example, a new color alternative to classical zircon palette can be obtained from the perovskite, a new and versatile lattice for ceramic pigments not considered on CPMA classification of industrial pigments. Neodymium perovskite pigments have been obtained by nonconventional methods [7–9]. Red, blue, green, and gray ceramic pigments based on FeNdO_3 perovskite (unmineralized sample or using $\text{BaF}_2+\text{MgF}_2$ flux agent for red and blue pigments, respectively) and CrNdO_3 (unmineralized sample or using

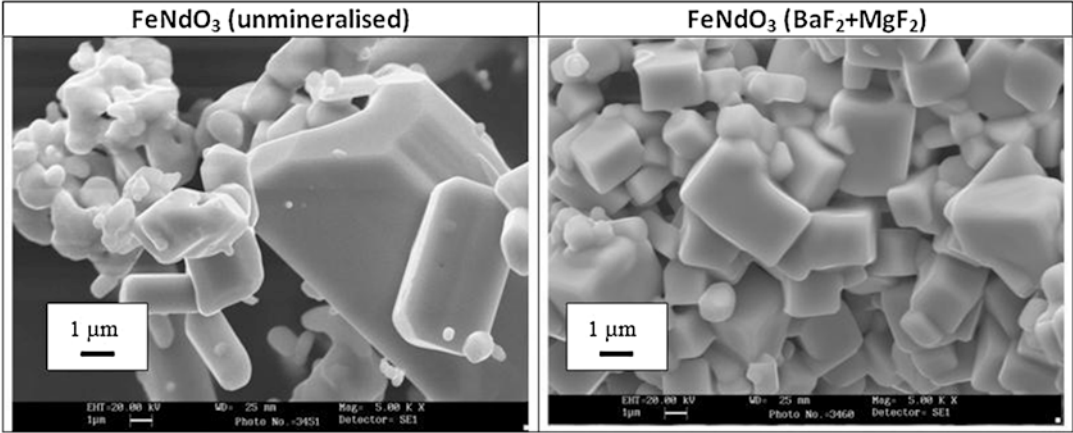
$\text{BaF}_2+\text{MgF}_2$ flux agent for green and gray pigments, respectively) can be obtained by ammonia coprecipitation from a solution of $\text{NdCl}_3 \cdot 6\text{H}_2\text{O}$, $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, $\text{CrCl}_3 \cdot 6\text{H}_2\text{O}$ salts in water media.

Dried powders were successively fired at 1,000, 1,100, and 1,200 °C during 6 h of soaking time, and the obtained powders were 5% enameled in a conventional glaze (1,050 °C). Characterizations of samples are shown in Figs. 3 and 4:

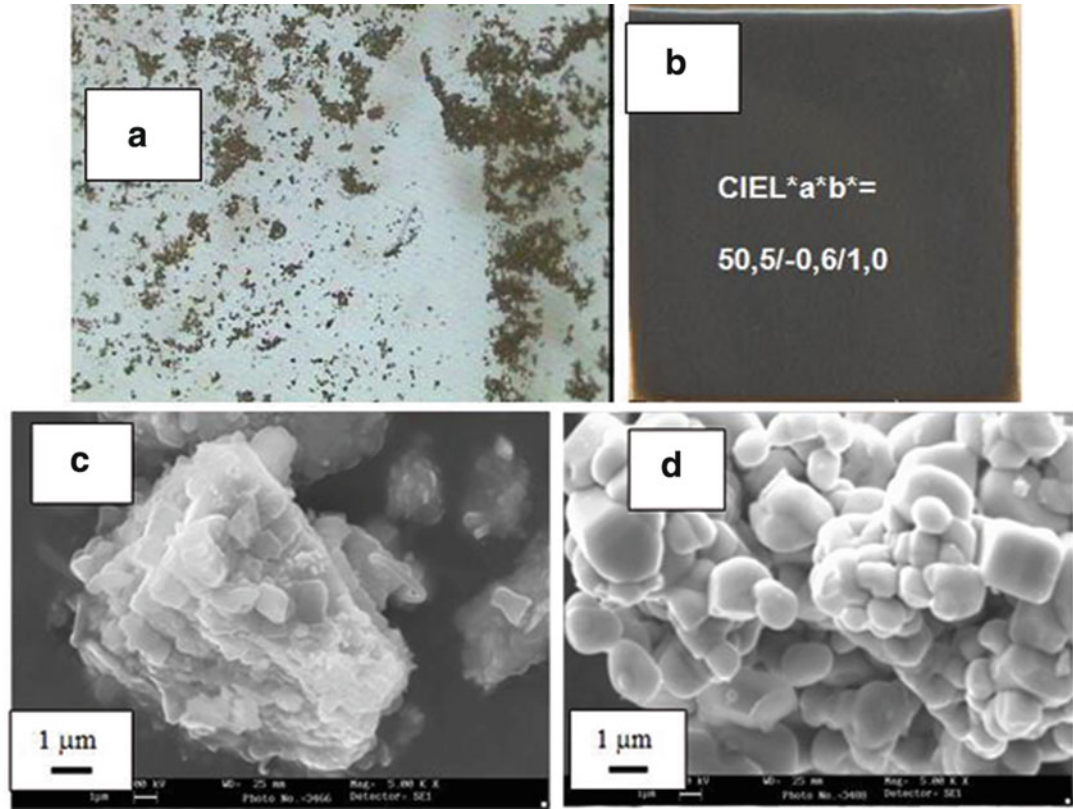
- (a) Powder optical lens view ($\times 40$) (Fig. 3a) shows red, blue, green, and gray color of the respective perovskites.
- (b) XRD (X-Ray Diffraction) analysis carried out on a diffractometer, using Cu K_α radiation, 20–70 °2 θ range, scan rate 0.05 °2 θ /s, 10 s per step, and 40 kV and 20 mA conditions, shows that perovskite is the only crystalline phase detected in CrNdO_3 samples, but residual Nd_2O_3 peaks are always observed on FeNdO_3 samples (Fig. 3b).
- (c) 5 w.% enameled samples in a conventional glaze (1,050 °C) show the visual red, blue, green, and gray color of the respective perovskite color associated to CIEL*a*b* color measurements (Fig. 3c), which were measured following the CIE (Commission Internationale de l'Eclairage) colorimetric method using a spectrophotometer, with standard lighting D65 (natural daylight) and standard observer of 10°. On this method, L* is a measure of brightness (100 = white, 0 = black) and a* and b* of chroma (–a* = green, +a* = red, –b* = blue, +b* = yellow).
- (d) UV-Vis-NIR (Ultraviolet-Visible-Near Infrared) spectroscopy of enameled samples, collected using a spectrometer through diffuse reflectance technique, shows bands associated to Fe^{3+} in octahedral coordination (in FeNdO_3) or Cr^{3+} in octahedral coordination (in CrNdO_3), respectively, along with sharp bands associated to Nd^{3+} f-f transitions [12]. Samples with $\text{BaF}_2+\text{MgF}_2$ flux agent addition show similar bands to unmineralized samples, but the absorbance level increases in the Vis-NIR range producing dark colors (dark blue and black color, respectively, on Fig. 3c).



Pigment, Ceramic, Fig. 3 Samples fired at 1,200 °C: (a) powder optical lens view (×40), (b) DRX, crystalline phases. p (perovskite MNdO₃, M = Fe or Cr), o(neodymium oxide Nd₂O₃), (c) enameled samples, (d) UV-Vis-NIR spectroscopy



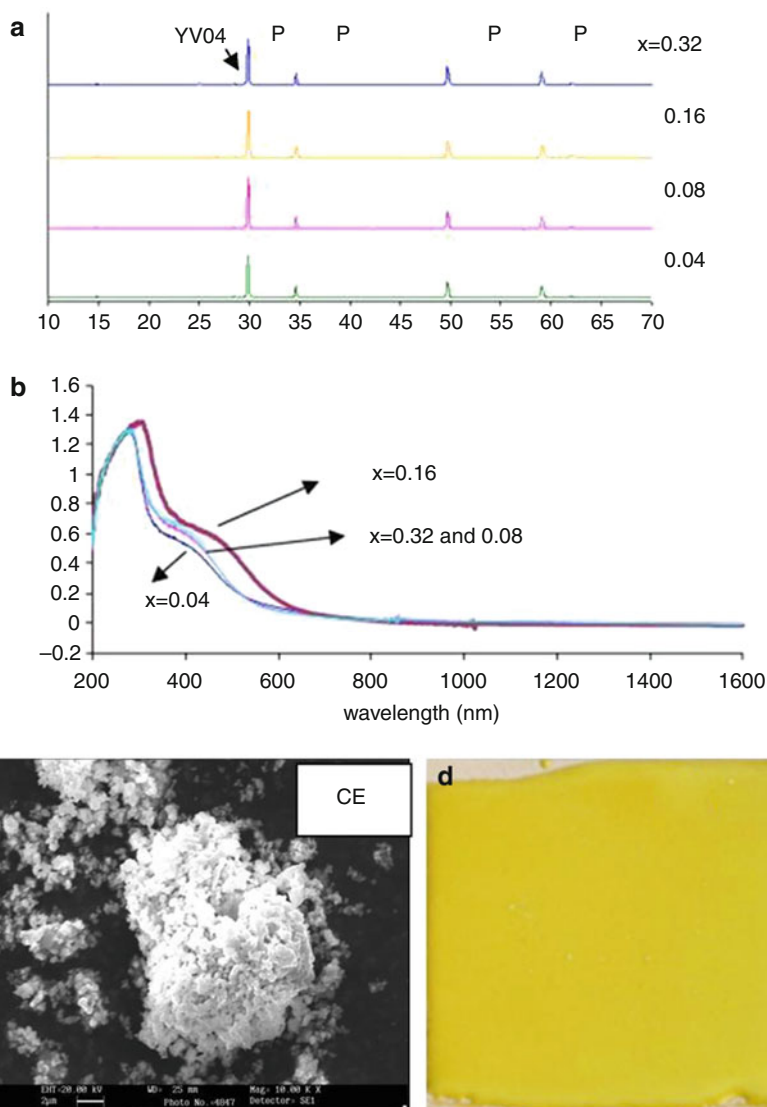
Pigment, Ceramic, Fig. 4 SEM micrographs of samples fired at 1,200 °C



Pigment, Ceramic, Fig. 5 CrNdO_3 fired at 1,100 °C: (a) powder optical lens view ($\times 40$), (b) enameled sample, (c) SEM micrograph of powder (500 °C), (d) SEM micrograph of powder (1,100 °C)

Pigment, Ceramic,

Fig. 6 (a) XRD diffractograms of $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_{7.0,14}\text{Na}_2\text{SiF}_6$ powders, CRYSTALLINE PHASES: P = $\text{Y}_2\text{Sn}_2\text{O}_7$, S = SnO_2 , Y = Y_2O_3 . (b) UV-Vis-NIR spectra of $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_{7.0,14}\text{Na}_2\text{SiF}_6$ powders, (c) SEM micrograph of $x = 0.16$ sample, (d) 5w% $x = 0.16$ glazed sample



The micrographs of samples of unmineralized and $\text{BaF}_2+\text{MgF}_2$ mineralized samples fired at $1,200^\circ\text{C}$, carried out by scanning electron microscopy (SEM), are shown in Fig. 4. Both samples show higher cubic crystalline morphology, but unmineralized sample shows higher size and heterogeneous particles that show low aggregation; in contrast $\text{BaF}_2+\text{MgF}_2$ mineralized sample shows a bimodal distribution of particle size (small $0.5\text{ }\mu\text{m}$ edge cubic particles and big $1.5\text{ }\mu\text{m}$ particles).

The preparation method of ceramic pigment allows obtaining new and best ceramic pigments.

There are different new methods of preparing ceramic pigments that produce new microstructures and solid solutions [1, 8].

Chemical methods as alternative to solid oxide reaction can be illustrated with above-discussed CrNdO_3 perovskite. A polyol route was used from $\text{NdCl}_3 \cdot 6\text{H}_2\text{O}$ and $\text{CrCl}_3 \cdot 6\text{H}_2\text{O}$: salts were dissolved in 200 ml of DEG (diethylene glycol) to prepare 5 g. of final product, then refluxed at 150°C during 2 h, and then charred at $500^\circ\text{C}/1\text{ h}$ and fired at $1,100^\circ\text{C}$ with soaking time of 6 h. SEM micrograph of powders charred at 500°C shows aggregates of incipient cubes (Fig. 5c),

well developed in fired sample at 1,100 °C (Fig. 5d). Powder develops, 5 w.% enameled in conventional ceramic glaze (1,080 °C), good black shades ($L^*a^*b^* = 50.5/-0.6/1.0$) without mineralizer addition (compare Figs. 5b and 3c).

Therefore above neodymiums could be an example of alternative to cyan, pink, and black pigments to the classical zircon color palette. Finally an alternative to yellow of praseodymium in zircon could be a $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_7$ yellow pigment based on pyrochlore crystal structure (Fd-3m) which describes materials of the types $\text{A}_2\text{B}_2\text{O}_6$ and $\text{A}_2\text{B}_2\text{O}_7$, where the A and B species are generally rare-earth or transition metal species, e.g., $\text{Y}_2\text{Ti}_2\text{O}_7$ [9]. A yellow pigment with the pyrochlore structure $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Ti}_{2-x}\text{O}_7$ is known since 1993 as a substitute for the decreasing variety of available yellow ceramic pigments due to the severe regulation of toxic lead and cadmium. The solubility limit of vanadium in this pigment was found to be 1.5 w.% as V_2O_5 or 0.13 as x in the above formula expression. Characterization of vanadium in the vanadium pyrochlore yellow pigment by electron spectroscopy by chemical analysis and electron spin resonance showed that the oxidation state of vanadium was V^{5+} and its yellow color mostly originated from V^{5+} substituted for Ti^{4+} [10, 11].

XRD (X-Ray Diffraction) results (Fig. 6a) indicate the crystallization of $\text{Y}_2\text{Sn}_2\text{O}_7$ as the only crystalline phase in all samples except in $x = 0.32$ that shows very weak peaks associated to YVO_4 along with perovskite peaks. Spectral and color rendering of samples by UV-Vis-NIR reflectance diffuse spectroscopy (Fig. 6b) shows a sharp band centered at 250 nm associated to the $\text{Sn}^{4+}-\text{O}^{2-}$ band transfer and an additional band centered at 500 nm responsible of yellow color and associated to $\text{V}^{5+}-\text{O}^{2-}$ band transfer. The maximum of absorbance is observed for $x = 0.16$ sample. CIEL*a*b* results summarized in Table 1 indicate that the yellow b^* parameter increases with x until $x = 0.16$ and decreases for $x = 0.32$ indicating a solubility limit of vanadium in this pigment around $x = 0.16$. The enameled powders were 5 w%. glazed in a CaO-ZnO-SiO_2 (1,080 °C) single firing glaze but do not produce color; in $\text{Na}_2\text{O-CaO-PbO-SiO}_2$ (1,000 °C) double

Pigment, Ceramic, Table 1 CIEL a*b* colorimetric coordinates for the $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_7$ yellow pigments

Sample (x)	L*	a*	b*
0.04	84.3	1.2	26.6
0.08	87.2	0.2	30.8
0.16	85.7	1.1	34.5
0.32	85.8	0.8	33.5

firing glaze, the color of all samples is similar (CIEL*a*b* = 78/5/35) (Fig. 6d).

In order to increase the reactivity of the system, the ceramic optimized composition $x = 0.16$ was prepared by ammonia coprecipitation (CO) and Pechini route (CI). In both routes Y ($\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$, $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$, and $\text{VOSO}_4 \cdot 8\text{H}_2\text{O}$ (ALDRICH) were used as precursors. In the CO route the precursors were solved in water (250 ml for 5 g of final product), then concentrated ammonia was dropped at 70 °C in continuous stirring until pH = 8; finally powder was obtained by drying in oven at 110 °C. In CI route, a molar ratio metallic cations/ethylene glycol/citric acid = 1:1:1 was used. Firstly precursors were solved in water (250 ml for 5 g of final product) and citric acid was solved at 70 °C in vigorous stirring, and then ethylene glycol was added and maintained 8 h for esterification. The obtained ester was dried at 110 °C and submitted to charring treatment at 250 °C. Na_2SiF_6 flux agent was added to CO or CI powders by manual mixture in an agatha mortar using acetone media. XRD results obtained at 1,000 °C/6 h for both CI and CE show peaks associated to pyrochlore along with unreacted SnO_2 and Y_2O_3 peaks; CO sample only shows weak peaks associated to SnO_2 along with pyrochlore. At 1,200 °C all samples show pyrochlore as the only crystalline phase detected. CO at 1,200 °C shows the best pigmenting results in double firing glaze ($L^*a^*b^* = 80.0/5.0/43.7$) than CE (79.3/2.6/35.4) and CI (81.7/0.1/31.8).

SEM micrographs of powders fired at 1,200 °C indicate the presence of aggregates of sub-micrometric particles in all samples (Fig. 6c). Size of aggregates in CE and CO sample (2–6 μm) is similar but more compact in CO sample and higher than in citrate powder

(1–4 μm). Also size of particles forming aggregates is higher for CE (500 nm) than for CO (300 nm) and higher than CI powder (150 nm). Likewise SEM-EDX (Scanning Electron Microscopy-Energy Dispersive X-Ray Spectroscopy) microstructural analysis of powders fired at 1,200 °C (not shown) indicates that codopants, vanadium, and calcium present a homogeneous distribution on samples. However the global content of calcium is higher than vanadium, probably due to vanadium evaporation during firing. Likewise the content of vanadium is slightly higher in CO and CE sample than in CI sample in agreement with color intensity observed on glazed samples.

Environmental Safety: Ecotoxicity Evaluation

The environmental safety using ceramic pigments is an important limitation that gives the severe regulation of toxic components such as lead or cadmium. It is necessary to check the safety of the pigments. The leaching test is usually used for determining the ecotoxicity of solids. For example, the European leaching test was performed at a ratio of liquid to solid sample $L/S = 16 \text{ L/kg}$. A mixture of 10 g solid sample and 160 mL distilled water was combined in borosilicate glass bottles and then agitated for 24 h at 10 rpm with a rotary agitator. The leachate was then collected by filtration through a 0.45 μm membrane filter. The toxicity of the leachates is evaluated using a battery of bioassays including the photobacterium *Vibrio fischeri* or *Photobacter phosphoreum* (Microtox test) and the crustaceans *Daphnia magna*. Microtox test was based on the bioluminescence measurement of the marine bacteria *V. fischeri*, within exposure time to leachates of 15 min., using the DeltaTox PS1 Analyzer.

The Microtox ecotoxicity test results for the above-discussed pigments are shown in Tables 2 and 3. In the case of the yellow of the pyrochlore $\text{Ca}_{0.16}\text{Y}_{1.84}\text{V}_{0.16}\text{Sn}_{1.84}\text{O}_7$, both unmineralized ceramic CE and sonocoprecipitated SONO powders fired at 1,200 °C/6 h are compared in Table 2.

Pigment, Ceramic, Table 2 EC50 Microtox results for pyrochlore of tin pigments

Sample	EC50 (ppm)
$\text{Ca}_{0.16}\text{Y}_{1.84}\text{V}_{0.16}\text{Sn}_{1.84}\text{O}_7$ CE 1,200 °C/6 h)	296,400
$\text{Ca}_{0.16}\text{Y}_{1.84}\text{V}_{0.16}\text{Sn}_{1.84}\text{O}_7$ SONO 1,200 °C/6 h	65,639

Pigment, Ceramic, Table 3 EC50 Microtox results for perovskite of neodymium pigments

Sample	EC50 (ppm)
FeNdO₃	86,664
2 %BaF ₂ +8%MgF ₂	
SONO fired at 1,100 °C/3 h	
CrNdO₃	54,398
2 %BaF ₂ +8%MgF ₂	
SONO fired at 1,100 °C/3 h	
(Mg_{0.5}Fe_{0.5})(Co_{0.5}Fe_{1.4}Cr_{0.1})O₄	2,336
SONO fired at 1,000 °C/3 h	

In the SONO method, the chloride precursors were dissolved in water/diethylene glycol = 1:1 v/v media and were coprecipitated in ultrasonic bath by dropping 12% ammonia solution until $\text{pH} = 8$. In agreement to European laws, the limit of EC50 (equivalent concentration of leached that decreases the bioluminescence of *V. fischeri* to 50%) is 3,000 ppm. Therefore the pigment accomplishes the European regulations.

In the case of the yellow of the neodymiums above described (violet of FeNdO₃ with 2% BaF₂+8%MgF₂ SONO fired at 1,100 °C/3 h and black of CrNdO₃ with 2%BaF₂+8%MgF₂SONO fired at 1,100 °C/3 h), the Microtox results are compared with the reference black of spinel (Mg_{0.5}Fe_{0.5})(Co_{0.5}Fe_{1.4}Cr_{0.1})O₄ SONO fired at 1,000 °C/3 h) in Table 3. In agreement with European limit of EC50 the perovskite of neodymium pigments accomplish the European regulations but not the spinel black due to released chromium. The pigments based on rare-earth components usually give eco-friendly pigments [12].

Cross-References

- CIELAB
- Color Vision Testing
- Colorant, Natural
- Dye, Ink-Jet

Acknowledgments Authors acknowledge the financial support given by FUNDACION CAJA CASTELLÓN-UJI, P1-1B2010-09 project.

References

1. (a) Monrós, G., Badenes, J.A., García, A., Tena, M.A.: El color de la Cerámica. Nuevos mecanismos en pigmentos para los nuevos procesados de la industria cerámica, pp. 146–179. Edited by Universitat Jaume I, ISBN 84-8021-449-X (2003). (b) R. A. Eppler Selecting Ceramic Pigments. Am. Ceram. Soc. Bull. **66**(11), 1600–1604 (1987) (c) Bucher, W., Schliebs, R., Winter, G., Buchelin, K.H.: In: Industrial Inorganic Chemistry, VCH, Berlin (1989)
2. (a) CPMA: Classification and chemical description of the complex inorganic color pigments, 4th edn. Dry Color Manufacturers Association, Alexandria (2010) (b) DCMA: Safe Handling of Color Pigments, Color Pigments Manufacturers Association Inc., 300 North Washington Street, Suite 102, Alexandria (1993)
3. Seabright, C.A.: Ceramic Pigments, US Patent., 2, 407–441 (1948)
4. (a) Monrós, G.: Soluciones sólidas V-ZrSiO₄, Tesis Doctoral, Universidad de Valencia (1991) (b) Lahuerta, J.: Colorantes Cerámicos Rojos Basados en Fe₂O₃, Tesis Doctoral, Universidad de Valencia (1993)
5. (a) Badenes, J.: Estudio de los sistemas Pr-ZrSiO₄ y Pr-ZrO₂, Tesis Doctoral, Universidad Jaume I de Castellón (2000) (b) Calbo, J.: Desarrollo de ecopigmentos negros de espinela dopada alternativos a los ferritos tradicionales mediante presión y procesados sol-gel, Tesis Doctoral, Universidad Jaume I de Castellón (2003)
6. Hopper, R.: The Ceramic Spectrum, 2nd ed. Chilton Book Co, ISBN 978-1-57498-302-9, Radnor, Penna (1984)
7. García, A.: Pigmentos cerámicos CMYK de base nanoestructurada obtenidos por sonocoprecipitación y ruta poliol, Tesis Doctoral, Universidad Jaume I de Castellón (2011)
8. García, A., Llusar, M., Sorli, S., Tena, M.A., Calbo, J., Monrós, G.: Eu³⁺-Nd₂O₃ blue pigments solid solutions. Brit. Ceram. Trans. **101**, 242–246 (2003)
9. Gargori, C., Galindo, R., Cerro, S., García, A., Llusar, M., Monrós, G.: Synthesis of a new Ca_xY_{2-x}V_xSn_{2-x}O₇ yellow pigment. Phys. Procedia. **8**, 84–87 (2010)
10. Subramanian, M.A., Aravamudan, G., Subba Rao, G.V.: Oxide pyrochlores: a review. Prog. Solid State Chem. **15**, 55–143 (1983)
11. Ishida, S., Ren, F., Takeuchi, N.: New yellow ceramic pigment based on codoping pyrochlore-type Y₂Ti₂O₇ with V⁵⁺ and Ca²⁺. J Am Ceram. Soc. **76**(10), 2644–2648 (1993)
12. García, A., Llusar, M., Calbo, J., Tena, M.A., Monrós, G.: Low toxicity red ceramic pigments for porcelanised stoneware from lanthanide-cerianite solid solutions. Green Chem. **3**, 238–242 (2001)

Pigments

- Colorant, Natural

Pigments, Organic

Robert Christie

School of Textiles & Design, Heriot-watt
University Scottish Borders Campus, Galashiels,
UK

Definition

A pigment is a finely divided solid, crystalline coloring material, manufactured in a finely divided form, and is essentially insoluble in its application medium [1, 2]. Pigments are classified chemically as either organic or inorganic [1–3]. The requirement for organic pigments to be insoluble distinguishes them from organic dyes which are generally required to be soluble, especially in water for textile coloration. Organic pigments are used mainly in the coloration of paints (coatings), printing inks and plastics and, to a certain extent, in a wider range of substrates, including paper, textiles, rubber and cosmetics. Their application generally involves incorporation into a liquid medium, for example a wet paint, ink or a molten thermoplastic, using a dispersion process in which clusters of pigment particles are broken down into primary

Robert Christie has retired.

particles and small aggregates. The pigmented medium then solidifies, by physical solidification, solvent evaporation or polymerization, and the finely divided pigment particles become fixed mechanically in the solid matrix.

Properties

Optical Properties

Organic pigments are used to modify the optical properties of a material into which it is incorporated. The most obvious effect is to provide the desired color. The color provided by an organic pigment is primarily dependent on its molecular structure. However, two further factors influence their color since they are used as solid, crystalline particles: the crystal structure, i.e., the way in which the molecules pack in their crystal lattice, and the particle structure, notably particle size and shape distribution [4]. Organic pigments generally provide bright, intense colors and in this sense offer advantage over their inorganic counterparts. Their color strength increases with a decrease in particle size, so for optimum color strength they are commonly manufactured at as small a particle size as possible, approaching the nanometer range.

However, providing color is not their only optical function. They may also be required to provide opacity. This feature is most important in paints which are commonly designed to obscure the surface to which they are applied. Alternatively, and in complete contrast, high transparency may be important. For example, multicolor printing generally uses inks of four colors, the three subtractive primaries, yellow, magenta and cyan, together with black, and transparency is essential to ensure that subsequently printed colors do not obscure the optical effect of the colors printed previously.

Fastness Properties (Technical Performance)

Organic pigments must be capable of withstanding the effects of their environment, both during processing and in their anticipated useful lifetimes. They are required to be fast to light, weathering (if external exposure is envisaged),

heat and chemicals (such as acids and alkalis) to a degree dependent on the demands of the application. Pigments are also required to show solvent resistance, which refers to their ability to resist dissolving in solvents that they may contact in their application, in order to minimize issues such as 'bleeding' and migration of color. Heat stability is critically important in pigments for applications which require processing at high temperatures and is thus a vital factor in the coloration of thermoplastics. The level of thermal stability required depends on the processing temperature for the polymer in question and on the time of exposure.

Dispersibility

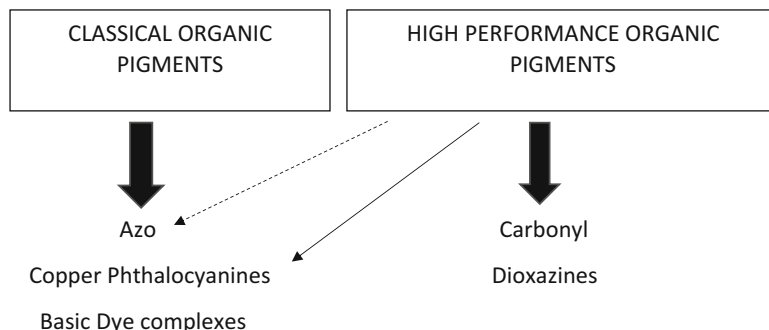
It is essential that an organic pigment is well dispersed into its application medium, for example during its introduction into a paint, printing ink or plastic article. Pigment powders consist of clusters of aggregated particles which are subjected to conditions which break down large aggregates into smaller aggregates and primary particles to the degree required to give the optimum properties. *Dispersibility* refers to the ease with which the desired level of dispersion is achieved.

Classification of Organic Pigments

Classification According to Technical Performance

Organic pigments may be classified broadly as classical or high performance (Fig. 1). Organic pigments were developed in the late nineteenth century as synthetic textile dyes became established. Many early organic pigments were based on water-soluble dyes rendered insoluble by precipitation onto colorless inorganic substrates, such as alumina and barium sulfate. These products were referred to as *lakes*. A significant early development was the introduction of a range of azo pigments, which provide the basis for the most important yellow, orange and red organic pigments currently in use. These so-called classical azo pigments offer bright, intense colors, although only moderate

Pigments, Organic,
Fig. 1 Organic Pigment
Classification



performance in terms of fastness properties. A critical historical event was the discovery, in 1928, of copper phthalocyanine. This blue pigment was the first product to offer the outstanding intensity and brightness of color that is typical of organic pigments, combined with an excellent set of fastness properties. The discovery stimulated the quest for other chemical types of organic pigment which could emulate the high performance properties of copper phthalocyanine in the yellow, orange, red and violet shade areas, an activity that gained further impetus from the growth of applications that demand high levels of technical performance, such as automotive paints and the plastics and synthetic fiber industries [4].

Classification According to Chemical Structure

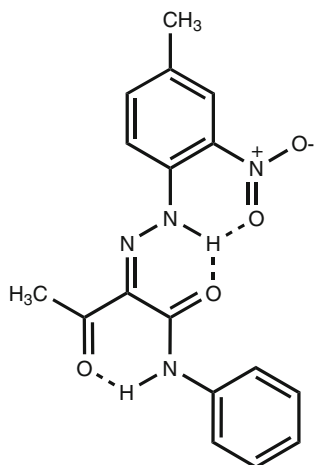
The broad chemical classes encountered in industrial organic pigments are shown in Fig. 1, which also illustrates their relationships with the classical and high-performance classifications.

Product Overview

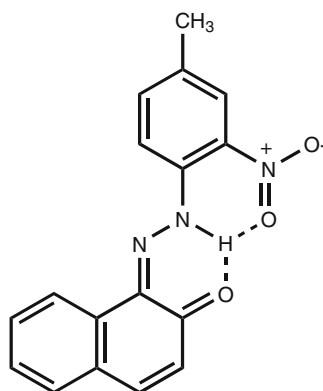
This section provides an overview of the characteristic structural features of the important chemical types of commercial organic pigments, together with a discussion of how these structural features determine their suitability for particular applications. Structure-property relationships in organic pigments are complicated by the dependence of color and technical performance not only on the molecular structure but also on the solid-state crystal structural arrangement and on the size

and shape distribution of the pigment particles [5]. Classical organic pigments are commonly prepared in as fine a particle size as is technically feasible in order to give maximum color strength and brightness and transparency, a combination that means that organic pigments are especially well suited to printing ink applications, where the particle size is commonly below 0.05 μm . For some applications, notably surface coatings, a larger particle size is desirable for improved opacity and fastness properties, but this is achieved at the expense of color strength. Organic pigments are unable to provide the degree of opacity that is typical of inorganic pigments, because of the lower refractive index associated with organic crystals. With organic pigments, there is an optimum size for opacity, often in the range 0.2–0.4 μm .

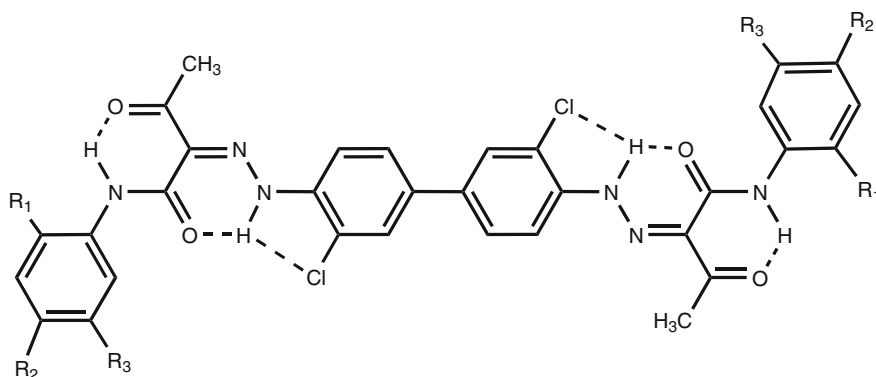
Commercial organic pigments exhibit a rather variable set of fastness properties, dependent on both the molecular structure and on the nature of the intermolecular association in the crystalline solid state. Since organic molecules commonly exhibit some tendency to dissolve in organic solvents, structural features are incorporated that enhance solvent resistance, by enhancing the strength of intermolecular interactions within the crystal structures. For example, an increase in the molecular size of the pigment generally improves solvent resistance. Hydrogen bonding is a highly important feature in organic pigment structures, promoting enhanced stability (e.g., to light and heat), and insolubility, as demonstrated in the numerous specific chemical structures presented in Figs. 2, 3, 4, 6 and 7 later in this section. Of particular importance in this respect is the



1:CI Pigment Yellow 1



2:CI Pigment Red 3

3a: $R_1 = R_2 = R_3 = H$

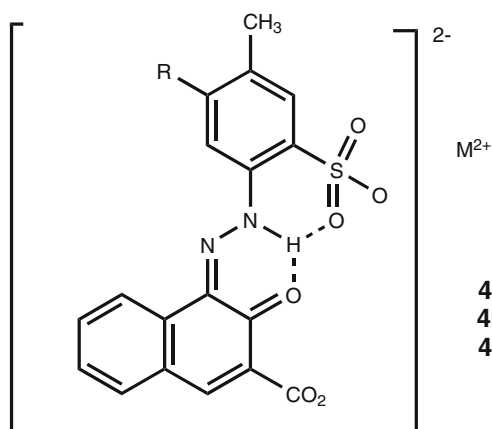
CI Pigment Yellow 12

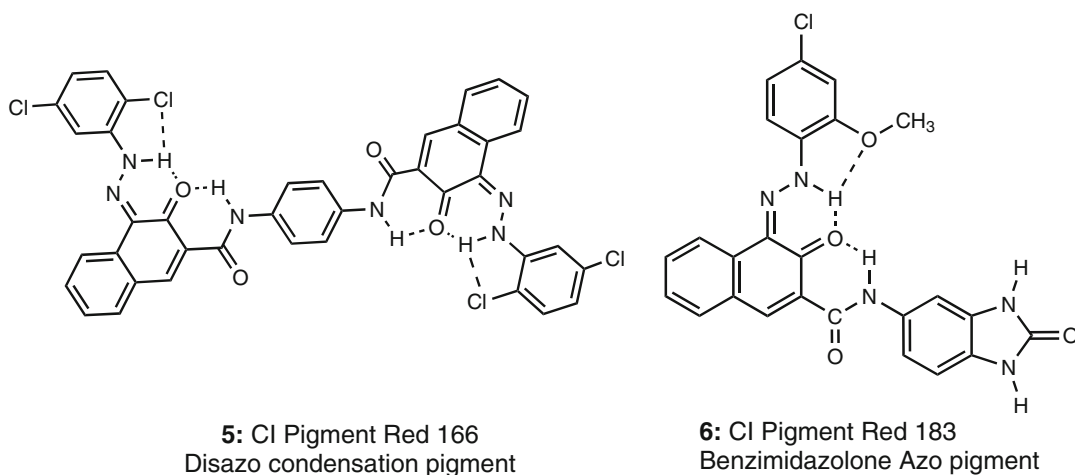
3b: $R_1 = R_2 = CH_3$; $R_3 = H$

CI Pigment Yellow 13

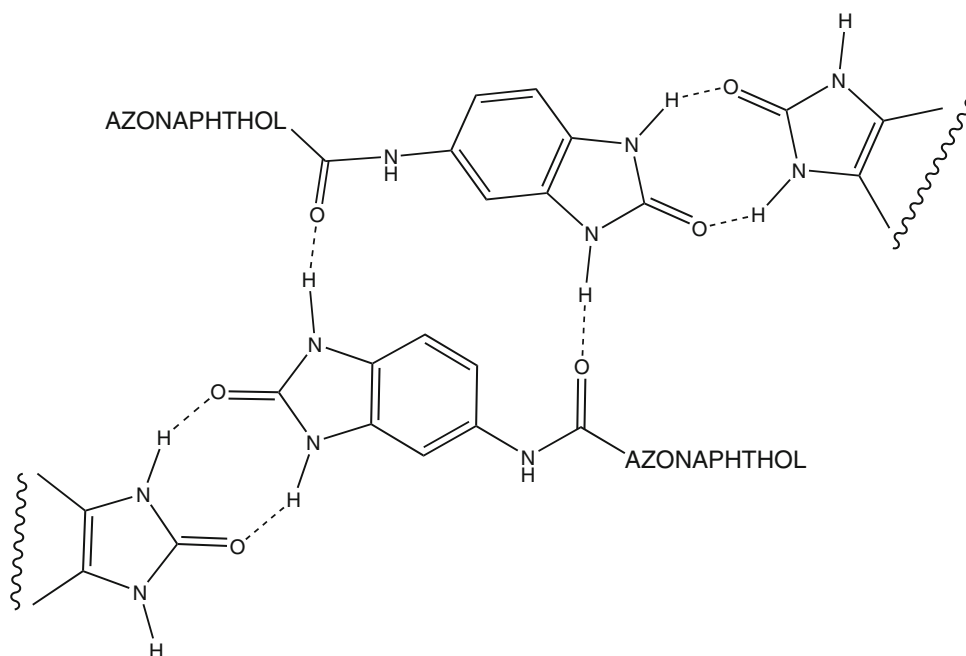
3c: $R_1 = R_3 = OCH_3$; $R_2 = Cl$

CI Pigment Yellow 83

4a: $R = H$; $M = Ca$ CI Pigment Red 57:14b: $R = Cl$; $M = Ca$ CI Pigment Red 48:24c: $R = Cl$; $M = Mn$ CI Pigment Red 48:4



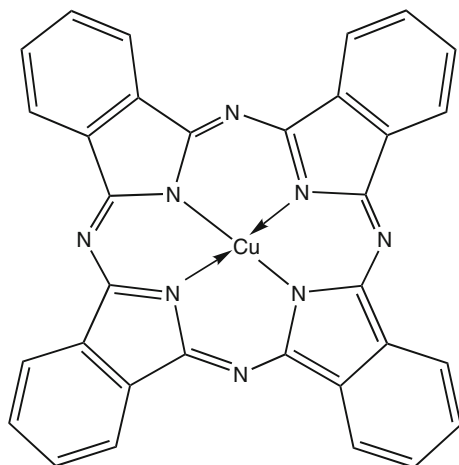
Pigments, Organic, Fig. 3 Examples of high-performance azo pigments



Pigments, Organic, Fig. 4 Intermolecular association in benzimidazolone azo pigments

amide (-NHCO-) group which features prominently in the chemical structures of organic pigments as a result of its ability to participate in strong hydrogen-bonding, both intramolecular and intermolecular. A related structural feature is found in certain pigments which contain N-H and

C=O groups in ring systems, an arrangement that leads to similar high technical performance as a result of strengthening the intermolecular association throughout the crystal structure. Examples of such arrangements can be found in Figs. 4 and 7. The incorporation of halogen substituents



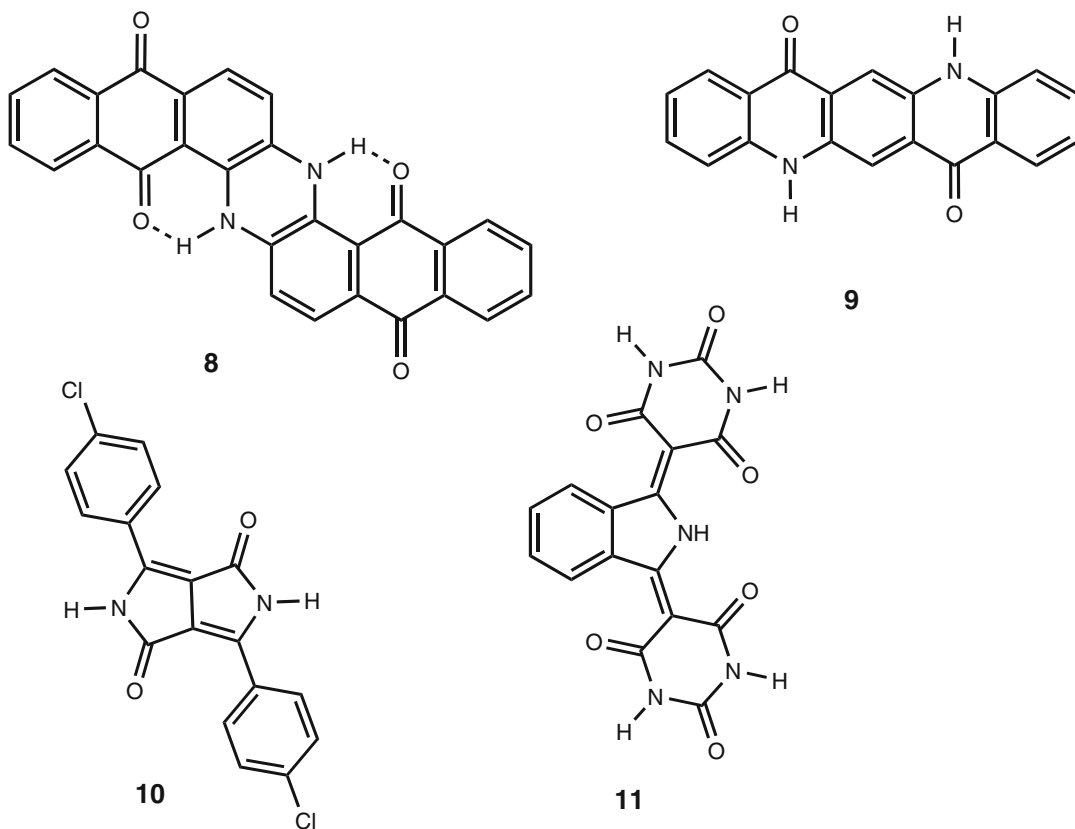
7:1 Pigment Blue 15 (CuPc)

Pigments, Organic, Fig. 5 The structure of copper phthalocyanine

and of metal ions, particularly alkaline earths and transition elements, can also have a beneficial effect on fastness properties.

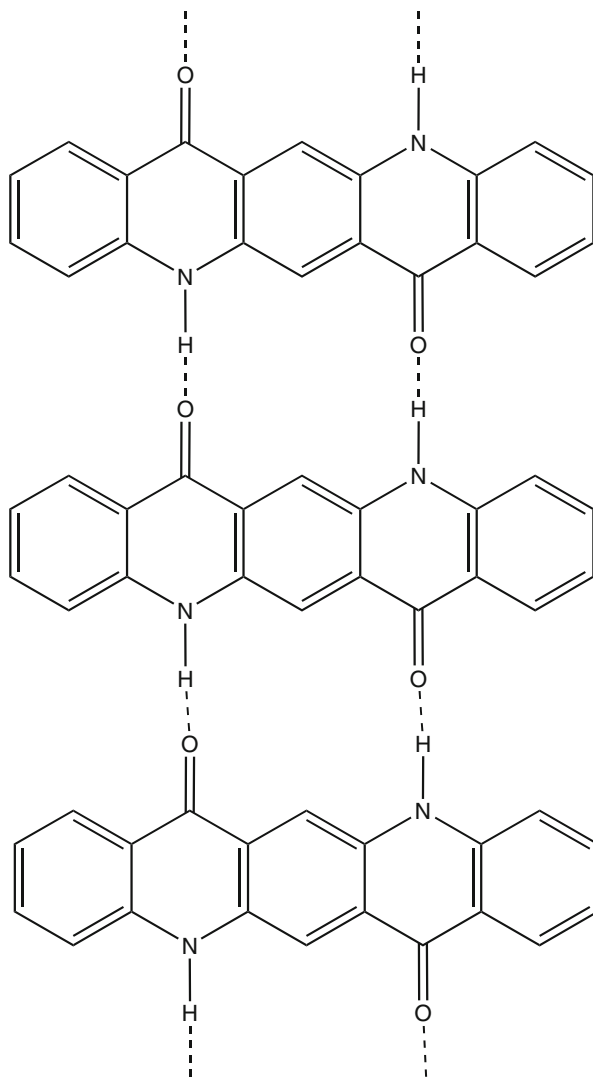
Azo Pigments

Azo pigments dominate the yellow, orange and red shade areas in commercial organic pigments [6]. The chemical structures of some important classical azo pigments are shown in Fig. 2. Many texts follow the convention to illustrate the structures in the azo tautomeric form. In this text, however, the structures are illustrated as ketohydrazones, since structural studies of industrial azo pigments have demonstrated that the pigments exist exclusively in this form in the solid state [1, 2, 5, 6]. Simple classical monoazo pigments such as Hansa Yellow G (**1**) and Tolidine Red (**2**) are products that provide bright colors and good lightfastness, but inferior solvent



Pigments, Organic, Fig. 6 The structures of selected carbonyl pigments

Pigments, Organic,
Fig. 7 Intermolecular
 association in the crystal
 structure of quinacridone (9)



resistance. The good lightfastness is attributed to the stability provided by the intramolecular hydrogen-bonding in six-membered rings. The inferior fastness to organic solvents is due to their small molecular size and intermolecular interactions in the crystal structures involving essentially only weak dipolar and van der Waals' forces. The use of these pigments is largely restricted to decorative paints since the pigments resist dissolving in the solvents used (water or aliphatic hydrocarbons) at the low temperatures involved in these applications. The most important yellow and orange azo pigments, particularly for printing inks are disazo pigments (diarylides),

exemplified by compounds **3a-3c**. They exhibit higher color strength and transparency than the corresponding monoazo pigments, properties particularly suitable for printing inks, and improved solvent fastness, attributable to the larger molecular size. The most important classical red azo pigments are metal salts, such as compounds **4a-4c**, derived from azo dyes containing $\text{-SO}_3^-\text{Na}^+$ or $\text{-CO}_2^-\text{Na}^+$ groups by replacement of the Na^+ ions with divalent metal ions, notably Ca^{2+} , Sr^{2+} , Ba^{2+} and Mn^{2+} [7]. They are products of high color strength and brightness, high transparency and good solvent resistance. CI Pigment Red 57:1, **4a**, is the pigment principally used to

provide magenta inks for multicolor printing. These products are complex structurally [8]. In the crystal structure, the metal atoms bond to carboxylate, keto and sulfonate oxygen atoms and to water molecules, with intermolecular bridging leading to a polymeric ladder structure.

There are two classes of high-performance azo pigments: disazo condensation pigments and benzimidazolone azo pigments. The chemical structures of representative examples of these products are illustrated in Fig. 3. Disazo condensation pigments constitute a range of durable yellow, red, violet and brown products, with structures such as compound **5** (CI Pigment Red 166). These pigments derive their name, from the elaborate synthetic procedures involved in their manufacture, which involves a condensation reaction. A second group of high-performance azo pigments contain the benzimidazolone group, exemplified by CI Pigment Red 183, **6**, which range in shades from yellow to bluish-red and brown and exhibit excellent fastness properties. Their high stability to light and heat and their insolubility is attributed to extensive intermolecular association as a result of hydrogen-bonding and dipolar forces in the crystal structure, as illustrated in Fig. 4.

Copper Phthalocyanines

Copper phthalocyanines provide by far the most important of all blue and green pigments [9]. Copper phthalocyanine, **7** (CuPc, CI Pigment Blue 15; Fig. 5), is arguably the single most important organic pigment. Although generally regarded as a classical organic pigment, it could equally well be described as a high-performance organic pigment. It finds widespread use in most pigment applications because of its brilliant blue color and its excellent resistance towards light, heat, solvents, acids and alkalis. In addition, copper phthalocyanine is relatively inexpensive as it is manufactured in high yield from low cost starting materials. CuPc exhibits polymorphism, referring to the ability to exist in different crystal forms. Both the reddish-blue α - and greenish-blue β -forms are of commercial importance. The β -form of the pigment is almost always the pigment of choice for cyan printing inks, and it is

suitable also for use in many paint and plastics applications. The most important green organic pigments are halogenated copper phthalocyanines, CI Pigment Green 7, in which the 16 ring hydrogen atoms of the CuPc molecule are replaced virtually completely by chlorine, and CI Pigment Green 36, a designation which incorporates a range of bromo- and bromochloro-copper phthalocyanines. The phthalocyanine greens exhibit the same outstanding color and technical performance as the blue pigments and find equally widespread use in the coloration of paints, printing inks and plastics.

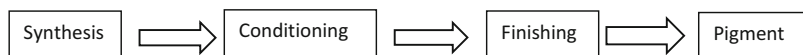
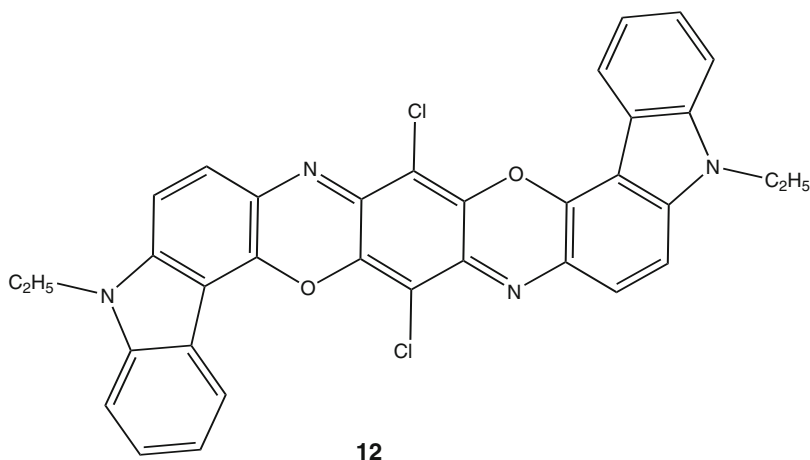
Cationic Dye Complex Pigments

These bright, intensely colored organic pigments are derived from selected basic (cationic) dyes rendered insoluble by associating with large counteranions, such as phosphomolybdotungstates. They are especially important as violet, blue and green pigments for printing inks.

Carbonyl Pigments

Carbonyl colorants (Fig. 6) contain the carbonyl (C=O) group as the essential structural component and chromophore [10]. A wide variety of structural types have been developed to provide yellow, orange, red and violet shades with the high-performance properties comparable with those of copper phthalocyanine blue and green pigments. A number of vat dyes for textile applications, after conversion to an appropriate pigmentary physical form, are suitable for use in many paint and plastics applications [10]. Examples of these so-called vat pigments include anthraquinones, such as Indanthrone Blue, **8** (CI Pigment Blue 60). Other high performance carbonyl pigments include quinacridone, **9** (CI Pigment Violet 19), diketopyrrolopyrrole (DPP) pigments, such as CI Pigment Red 254, **10**, and isoindoline CI Pigment Yellow 139, **11**. The excellent lightfastness, solvent resistance and thermal stability of carbonyl pigments may be explained in many cases by intermolecular association in the solid state as a result of a combination of hydrogen bonding and dipolar forces. A diagrammatic illustration of the intermolecular hydrogen bonding which exists in the crystal

Pigments, Organic,
Fig. 8 The chemical
 structure of dioxazine **12**



Pigments, Organic, Fig. 9 The stages of organic pigment manufacture

lattice arrangement of the quinacridone, **9** is given in Fig. 7 [11].

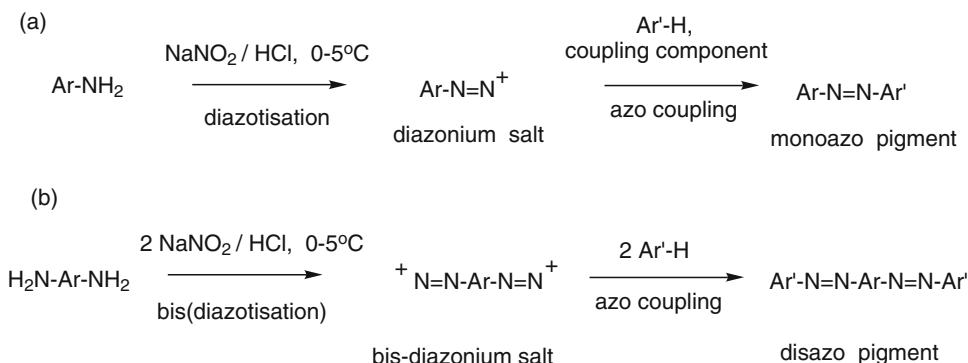
Dioxazines

The dioxazine, **12** (Carbazole Violet), whose polycyclic structure is shown in Fig. 8, is an important violet high-performance organic pigment.

Manufacture

The process by which a pigment is manufactured may be considered as involving a series of stages as shown in Fig. 9. The first of these is *synthesis*, the sequence of chemical reactions which form the pigment molecules. This is illustrated later for azo and copper phthalocyanine pigments. However, simply carrying out this chemistry does not necessarily give the product in a form that is optimum for a particular application. In a second phase, commonly referred to as *conditioning*, the product is developed into an appropriate crystalline form, with a controlled particle size and shape distribution, and the surfaces of the particles may be modified by suitable treatments. Depending on the pigment, these two stages may

be separate or combined into a single process. An important reason for surface treatment is to enhance the dispersibility of a pigment as it is introduced into a paint, printing ink or plastic article. In the dispersion process, clusters of aggregated particles are subjected to conditions which break down large aggregates into small aggregates and primary particles to the level required to give the optimum optical properties. The nature of the surfaces of pigment particles is of importance in their application performance because it is these surfaces that are in contact with the application medium. The particle surfaces formed during the synthesis of organic pigments often have highly polar character and may show poor compatibility with many of the non-polar (hydrophobic) organic media used in the applications. The pigments are therefore often treated with materials such as organic surface-active agents or resins to change the nature of the particle surfaces. In the treatment of the pigment with a surface-active agent, which contains separate hydrophilic and hydrophobic regions, hydrophilic groups attach to a polar pigment surface and hence the pigment effectively acquires a new surface which is hydrophobic and more compatible with an organic application medium. This



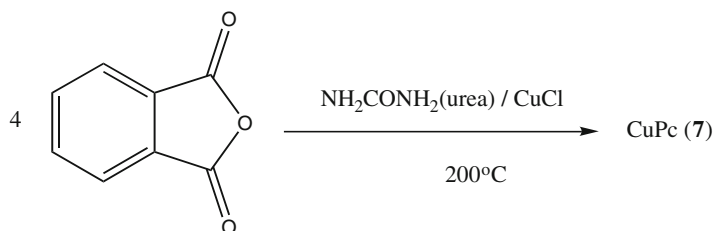
Pigments, Organic, Fig. 10 Synthesis of monoazo and (symmetrical) disazo pigments

surface also presents a barrier so that the pigment particles are less susceptible to aggregation, and those aggregates that are formed are more easily broken down by the dispersion process. Finally, the pigment is subjected to *finishing* whereby it is converted into a form acceptable to the user, for example by grinding to a fine powder or incorporating into a concentrated dispersion. The conditions of synthesis of those organic pigments prepared by precipitation from reaction in water are commonly controlled to ensure that they are formed directly in an appropriate particle size range. Commonly, they precipitate in a poorly-defined crystal form, and so are subsequently treated at elevated temperatures, for example in the presence or absence of solvents, water in the case of classical azo pigments or organic solvents in the case of high-performance pigments, in order to develop high crystallinity and ensure that the products show optimum optical properties. Some high-performance organic pigments, synthesized at high temperatures in organic solvents, are produced initially in a large particle size form, which must be reduced to a fine particle size to make them useful pigmentary products.

As illustrated in Fig. 10(a), monoazo pigments, such as products (1) and (2) are manufactured using the two-stage process of diazotization, by treatment of a primary aromatic amine with sodium nitrite in an aqueous acidic medium cooled with ice, followed by azo coupling of the resulting diazonium salt with an appropriate coupling component [6]. To ensure that the pigments are obtained in high yield and purity, it is essential

to maintain careful control of experimental conditions to minimize the formation of side products which might adversely affect the color and performance of the pigments. When azo coupling is complete, the aqueous suspension is usually heated to the boil for a period of time, often in the presence of surface-active agents, to refine the pigment particles and develop crystallinity. Diarylide pigments such as compounds (**3a-3c**) are derived synthetically from bis-diazotization of an aromatic diamine which is subsequently coupled with two moles of coupling component as shown in Fig. 10(b). Metal salt azo pigments, such as compounds (**4a-4c**) are prepared by traditional diazotization/coupling processes to form the sodium salt of an azo dye which is treated with a solution of a salt of the divalent metal to displace the sodium and form the insoluble pigment.

Although the structure of copper phthalocyanine is rather complex, its synthesis is remarkably straightforward. It may be prepared in virtually quantitative yield from readily available, low cost starting materials [9]. In the most important method, as illustrated in Fig. 11, phthalic anhydride is heated with urea, the source of nitrogen in the process, copper (I) chloride and a catalytic amount of ammonium molybdate at around 200 °C in a high boiling organic solvent. This process produces crude copper phthalocyanine as large particles, which require to be reduced in size, for example by milling in the presence of inorganic salts. Carbonyl and dioxazine high-performance pigments tend to have diverse and

Pigments, Organic,**Fig. 11** Synthesis of Copper phthalocyanine

individualized manufacturing routes, details of which can be found elsewhere [10].

Current and Future Developments in Organic Pigments: A Brief Overview

Recent decades have seen major changes in the organic pigment manufacturing industry. Manufacturing capacity has declined substantially in the traditional heartlands of Europe and the USA where the original discoveries took place and the products and processes were developed. This has been accompanied by rapid growth of the industry in Asia, mainly in China and India. The industry currently produces a mature range of products for conventional applications as the cost of introducing new products to the market has become increasingly prohibitive, not only in terms of R&D effort but also in addressing the demands of toxicological evaluation and environmental legislation. Opportunities have been presented by the emergence of a range of applications in new technologies, demanding new types of organic colorant referred to as functional dyes and pigments, or as π -functional materials. Organic pigments have been developed or modified for use in reprographic techniques, such as electrophotography and digital inkjet printing, and in electronic display technologies. Addressing environmental issues has become a high priority. It is reasonably reassuring that organic pigments themselves are generally considered to be of low toxicity because of their high insolubility. However, there are a few concerns about trace amounts of toxic by-products formed in their manufacture or application. For example, diarylide pigments are no longer recommended for high-temperature applications on the basis of

evidence that they may cleave thermally releasing the highly toxic starting material 3,3'-dichlorobenzidine (DCB) [12]. Certain azo pigments also have the potential to form polychlorobiphenyls (PCBs) during their manufacture [13].

References

1. Hunger, K., Schmidt, M.U.: Industrial Organic Pigments: Production, Crystal Structures, Properties, Applications, 4th edn. Wiley-VCH, Weinheim (2018)
2. Christie, R.M.: Pigments. In: Christie, R.M. (ed.) Colour Chemistry, 2nd edn, pp. 214–249. Royal Society of Chemistry, Cambridge (2015)
3. Buxbaum, G. (ed.): Industrial Inorganic Pigments, 2nd edn. Wiley-VCH, Weinheim (2008)
4. Faulkner, E.B., Schwarz, R.J. (eds.): High Performance Pigments, 2nd edn. Wiley-VCH, Weinheim (2009)
5. Hunger, K.: The effect of crystal structure on the colour application properties of organic pigments. Rev. Prog. Coloration. **29**, 71–84 (1999)
6. Christie, R.M.: Azo dyes & pigments. In: Christie, R.M. (ed.) Colour Chemistry, 2nd edn, pp. 72–98. Royal Society of Chemistry, Cambridge (2015)
7. Christie, R.M., Mackay, J.: Metal salt azo pigments. Color. Technol. **124**, 133–144 (2008)
8. Bekö, S.L., Hammer, S.M., Schmidt, M.U.: Crystal structures of the hydration states of pigment red 57:1. Angew. Chem. Int. Ed. **51**, 4735–4738 (2012)
9. Christie, R.M.: Phthalocyanines. In: Christie, R.M. (ed.) Colour Chemistry, 2nd edn, pp. 133–146. Royal Society of Chemistry, Cambridge (2015)
10. Christie, R.M.: Carbonyl dyes & pigments. In: Christie, R.M. (ed.) Colour Chemistry, 2nd edn, pp. 99–132. Royal Society of Chemistry, Cambridge (2015)
11. Paulus, E.F., Leusen, F.J.J., Schmidt, M.U.: Crystal structures of quinacridones. Cryst. Res. Technol. **9**, 131–143 (2007)
12. Az, R., Dewald, B., Schnaitmann, D.: Pigment decomposition in polymers in applications at elevated temperatures. Dyes Pigments. **15**, 1–14 (1991)
13. Rodenburg, L., Guo, L., Christie, R.M.: Polychlorinated biphenyls in pigments: inadvertent production and environmental significance. Color. Technol. **131**, 353–369 (2015)

Pixel Light

- [Automotive Front Lighting System \(Status 2020\)](#)

Planckian Locus

- [Blackbody and Blackbody Radiation](#)

Planckian Radiator

- [Blackbody and Blackbody Radiation](#)

Plasmonics

- [Surface Plasmons](#)

Polychromatism

- [Iridescence \(Goniochromism\)](#)

Polychromy

José Luis Caivano
Secretaria de Investigaciones FADU-UBA,
Universidad de Buenos Aires, and Conicet,
Buenos Aires, Argentina

Definition

Polychromy refers to the combination of many colors in a visual scene, whether it is a natural landscape or a man-made arrangement of things, for instance, a piece of art, a design, an object, a building, an urban landscape, etc. The word is

formed by the Greek words *poly* (many) and *chroma* (color).

Overview

Colors can be divided into two separate classes: *chromatic colors* (such as red, green, blue, yellow, and any other tone), and *achromatic colors* (such as white, black, and grays). The characteristic of chromatic colors is that they have a definite hue, in addition to a certain saturation and lightness. Achromatic colors do not have a definite hue, and their saturation has a zero value; they are only distinguished by their lightness. When many chromatic colors of different hue appear together, this is called a *polychromatic* composition. Generally, more than two (at least three) colors of different hue are needed to properly speak of polychromy. The usual terminology differentiates *monochromy* (one color), *dichromy* (two colors), and *polychromy* (more than two colors) [1]. But sometimes the terms *trichromy* and *tetrachromy* may be employed, and thus, in this case, polychromy is applied to more than four colors ([2], pp. 279–280). If a composition uses colors of different saturation and lightness but of the same hue (for instance, different kinds of yellow that could be lighter or darker and more or less saturated), then this arrangement is said to be a *monochromatic* color composition or a monochrome arrangement (*mono*: one; *chroma*: color). If a composition uses only gray colors of different lightness (which may also include black and white), it is called an *achromatic* composition.

While it may seem from the Greek etymology that the words *color* and *chroma* are equivalent, modern conceptions make a distinction between them. *Color* refers to a visual sensation in general, which includes not only the categories of red, green, blue, yellow, etc. but also gray, black, and white. The term *hue* is used more specifically to name the quality of a chromatic color that makes yellow look different to red, red to green, green to blue, etc. In opposition, gray colors, black, and white have no hue. In modern usage, *chroma* is considered a synonym of saturation, i.e., the strength, the colorfulness, or the degree of purity

of a certain hue. As previously said, gray colors, black, and white have no chroma (saturation), or have chroma zero. However, in order to avoid confusion with the original Greek etymological meaning (chroma = color), the term *saturation* is recommended, rather than *chroma*.

Then, it has to be taken into account that the word *color* is often used with two different semantic extensions: with a narrow sense, meaning “chromatic color,” and with a wide sense, including also grays, white, and black as color sensations. The narrow meaning is evident when some people say that black is not a color (because it is considered “the absence of color”), white is not a color (because it is “the sum of all colors”), and grays are not colors either. From a more comprehensive point of view, and in modern usage, this is a misconception.

All the previous clarifications are important in order to understand that *polychromy* means not just “many colors” but many chromatic colors of different hue. Polychrome compositions may include, of course, achromatic colors, but their presence must be accompanied by chromatic colors of different hue in order to be truly polychromatic.

Polychromy in Color Harmony

From the quantity of hues that take part in a color combination, Antal Nemcsics states five types of color harmonies ([2], pp. 274–280):

- *Monochrome harmonies*, with colors taken from the same hue section in a color order system.
- *Dichrome harmonies*, with colors taken from two sections of hue, preferably complementary hues.
- *Trichrome harmonies*, with colors taken from three different hues, which may be evenly spaced in the whole chromatic circle or grouped forming a domain in one sector of it.
- *Tetrachrome harmonies*, with colors taken according to regular series from four hue triangles, which may be in four well-separated domains in the chromatic circle, in only one

domain, in a group of three hues in one domain against one hue opposite to them, or in two complementary pairs.

- *Polychrome harmonies*, with colors belonging to more than four different hues, preferably making groups of hues in definite domains.

Polychromatic Compositions Along History

The use of polychromy in the visual arts, decoration, and architecture has a long history. The small book by Thomas Goodwin reviews some of the earlier examples and concentrates mainly around the Middle Ages and the religious architecture in England [3]. Goodwin starts the introduction with a definition and then refers to the practice of polychromy in Egypt, Persia, and other antique civilizations:

The art of decorating the interiors of buildings with colour and gilding—Polychromy, as it is sometimes called, — is one which has been practised in all ages and among all nations. Its universal use among all people, as well barbarous as refined, is a satisfactory proof that it is founded upon the best and most natural principles. ([3], p. 1)

The Greeks painted their temples and family dwellings with many vivid colors of different hue, such as red, yellow, and blue. The idea that Greek temples had just the appearance given by white marble has been superseded long time ago, even when for some people it is still a novelty. Greeks started making wooden temples, and of course they painted them not only for decorative reasons but also to protect the wood. This usage created a style, and centuries later in the Greek culture and artistic practice, when the wood was replaced by marble for building temples, the custom of using colors applied by paints remained. And then, Greeks painted over the marble, producing a polychromatic architecture. The painted colors faded with time, the pigments disappeared, and for centuries the buildings remained as ruins of white marble for the later generations.

By the middle of the nineteenth century, Jacques-Ignace Hittorff succeeded in conveying the attention toward the discovery, made

Polychromy, Fig. 1 Two Greek sculptures

(c.490 BC): An archer from the Temple of Aphaia in Aegina, and a rider from the Acropolis in Athens. Present condition of the sculptures (left) and the polychrome reconstruction, as shown in the exhibition *Gods in color*, see <https://buntegoetter.liebieghaus.de/en/>

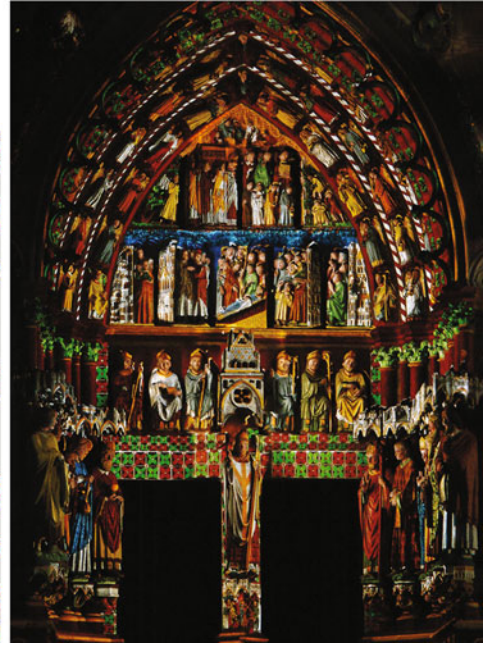
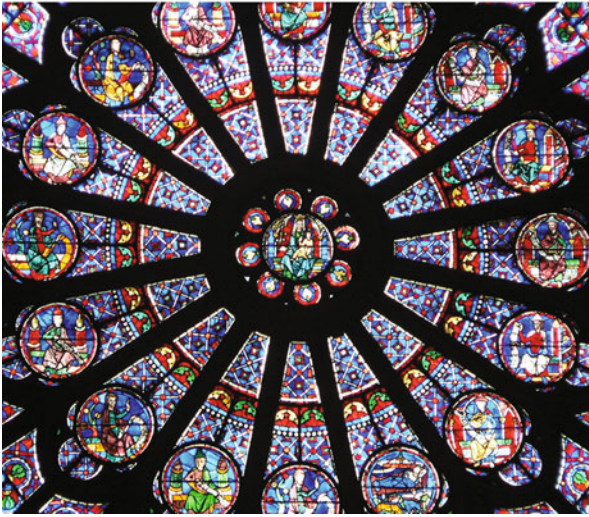


previously to him but neglected for many years, that the Greek architecture was not white – as it was believed for centuries based on the color of the ruins. It was polychromatic. Hittorff published these findings in his book of 1851 on the polychromy of Greek architecture, where he made a case of the temple of Empedocles in Selinus, Sicily [4]. The more general acceptance of this evidence came to change a long-held view about the Greek sense of beauty and harmony. However, it took some time to change the traditional notion, and the neoclassicist architects of the nineteenth century continued to make neoclassic buildings (based on the orders of Greek architecture) in gray, white, or with a monochromatic appearance. As a practicing architect, Hittorff was an exception, since he made polychromatic buildings [5].

Polychromy was also present in statuary in antique Greece. At the beginning of the nineteenth century, Quatremère de Quincy presented a detailed essay on the taste for polychrome sculpture among Greeks and Romans [6]. More recent

studies provide additional evidence that Greek artists used to paint with referential colors the marble statues, imitating the colors of skin, hair, representing the colors of clothes, etc. (Fig. 1). The analysis by Manzelli starts from the first studies on the polychromy of Greek and Roman sculptures, goes through the pigments and painting techniques of statues, and comes to consider the symbolic meanings of color in the ancient classical culture [7]. Since the turn of the third millennium, the research on the coloration of classical statues has been acquiring new interest, particularly with the exhibition at the Munich Glyptothek in 2003, and with the increasing number of publications on the subject during the past two decades [8].

Roman culture inherited the practice of polychromy from the Greeks and applied it also to statuary and architecture. In the present times, there is still physical evidence of the polychromy used in the interior of houses and buildings in the old Roman city of Pompeii, which remained



Polychromy, Fig. 2 Rose window in Notre Dame, Paris (left), and projected lights that recreate the original polychromatic appearance of the Amiens cathedral (right)

preserved for centuries under the ashes of the Vesuvius volcano.

There are evidences that the first Christians decorated the catacombs with color. This practice naturally continued with the first Paleochristian churches and during the Romanic style, and acquired a high degree of perfection in the Byzantine period, particularly with the mosaics. This technique of using polychromy has made it possible to preserve beautiful abstract decorations and figurative representations for more than a thousand years, up to the present days.

During the Middle Ages the polychromy in architecture was particularly prominent in the stained glass windows of the cathedrals. The colored windows also remained until now without deterioration. But, even when it is not always visible today, because pigments have faded, the interior of Gothic cathedrals exhibited architectural elements, such as columns and arches, painted in many colors, in addition to polychrome sculptures. What seems more surprising, from recent discoveries, is that not only the interior spaces but also the facades of many Gothic

cathedrals were painted with polychromatic combinations. A restoration of the western facade of the Amiens cathedral made in the 1990s revealed that it was originally painted in various colors and that by the thirteenth century it had a polychromatic appearance (Fig. 2).

In the Renaissance, most of the practice of polychromy applied directly on structural or decorative elements of architecture started to be neglected, and was partially replaced by representative frescoes and paintings on the walls and ceilings that endorsed architectonic spaces with color and images “borrowed” from the art of painting. According to Leon Solon, it was in this period that color began to be avoided and disassociated from architecture ([9], pp. 9–10). Even when the Renaissance implicated a classical revival, at this time, around the fifteenth century, very little or nothing was known about the polychromy of the Roman and Greek classical architecture, because the only remnants were achromatic or monochromatic ruins, and archeological studies had to wait until the nineteenth century to flourish. However, evidence of

polychromatic architecture of the medieval churches was supposedly present in the fifteenth century, but since the Renaissance was also a strong reaction against the Gothic style, this may have been another reason for abandoning color. Solon presents various other reasons and makes some speculations about this feature of Renaissance architecture ([9], pp. 11–15).

Owen Jones and Gottfried Semper were two nineteenth-century architects who deserve a special mention due to their contributions to the study of polychromatic architecture. In 1852, Jones published *An Attempt to Define the Principles which Should Regulate the Employment of Colour in the Decorative Arts*, in which he makes 22 propositions or rules for the use of color [10]. Jones was also responsible for coloring a Greek monument exhibited in the Crystal Palace, and he did this task based upon the researches of his time. In a book published in 1854, he explains and defends the criteria with which he selected and applied colors to the different parts of the monument [11]. An appendix of the book includes a fragment written by Gottfried Semper on the origin of polychromy, reviewing cultures such as Indians, Jews and Phoenicians, Greeks, Assyrians, Persians, Egyptians, Chinese, etc. Semper sets forth the hypothesis that this origin relies in textiles, which constitute the oldest manufactures used as covering, shelters, fences, or partitions of space in most cultures. Since textiles were usually dyed in many colors, it is supposed that this practice moved in a natural way to architecture, when textiles were replaced by strongest materials in order to build living spaces ([11], pp. 47–53). *The Grammar of Ornament*, published by Jones in 1856, constitutes an extraordinary source of historical color designs, systematically arranged and covering different geographies and periods, from the ancient times to the early seventeenth century [12].

Another source of information on polychromatic decoration, in this case concentrated on the medieval-style architecture, and particularly on the Gothic revival and Victorian style of the nineteenth century in England, is the book published in 1882 by the Scottish architects and brothers

George and William Audsley, with plenty of illustrations in the form of colored plates [13].

Architectural Polychromy in the Twentieth Century

The Case of Le Corbusier

Without any doubt, the richest and most complex personality of twentieth-century architecture has been Le Corbusier (Charles Edouard Jeanneret). His relationship with color has been complex too, and sometimes contradictory [5]. Between Bruno Taut, uncompromising defender and propagandist of polychrome architecture, and Walter Gropius, who theorizes about color but in his works uses it in extremely moderate doses, Le Corbusier appears to be midway. Precisely, he maintains some differences with Taut regarding color. About the chromatic audacity of Taut, Le Corbusier goes on to say, “My God, Taut is color blind!”

The early writings about color of Le Corbusier were published in the articles on purism and cubism in the journal *L'Esprit Nouveau*, in collaboration with the painter Amédée Ozenfant. An article from 1918 reads,

The idea of form precedes that of color. The form is preeminent, color is but one of its accessories. Color depends entirely of the material shape: the concept of sphere, for instance, precedes the concept of color; it is conceived as a colorless sphere, a colorless plane, color is not conceived independently of some support. Color is coordinated with form, but the reciprocal is not true. ([14], p. 42)

Some other articles published in 1921, 1923, and 1924 in the same journal proceed more or less in the same vein, that is, denying any importance that color might have in the construction of space in painting [15]. The curious fact is that a few years later, in his writings on architectural polychromy of 1931, Le Corbusier seems to have changed his mind completely, to the extent of quoting and agreeing with Fernand Léger, who said, “Man needs colors to live, it is an element as necessary as water and fire.” In addition, Le Corbusier describes examples of his own use of color in order to drastically change the spatial

perception of architecture, as in the neighborhood designed and built in Pessac [16].

In his monograph written for the exhibition of the *Pavillon des Temps Nouveaux*, Le Corbusier includes a chapter entitled “Polychromy = Joy,” in which he associates the creative ages of architecture with the vitality of chromatic color and relates the stagnant academicism to sad gray ([17], pp. 22–23). It seems that in both his theories and his works, Le Corbusier evolved toward a more conscious and thorough consideration of the power that color has to modify the spatial environment. This is especially evident in the buildings projected and built after World War II, in what is called his “brutalist” period.

As with the works by other architects of the modern movement, Le Corbusier’s architecture (which was reproduced in black-and-white photography in its time) is being reconsidered in the light of recent restorations and studies. The *Ville Savoye*, a paradigm of modern architecture of the 1920s and 1930s, was not white (as it appeared in many publications) but developed a complex polychromatic Scheme [18].

Le Corbusier’s text on architectural polychromy, written in the 1930s, remained as an unpublished manuscript. The first edition of this text, including also the color collections produced for the Salubra wallpaper company, was edited by Arthur Rüegg and published in 1997 [16, 19]. As for the polychromy in the purist architecture of Le

Corbusier, and his evolution, see also the book by Jan De Heer [20].

Polychromy in Modern Architecture and Beyond

Apart from the particular case of Le Corbusier, there are many other examples of polychromy in architecture in the twentieth century. At the beginning of the century, the Art Nouveau and Art Deco styles were particularly profuse with colors. The Neoplasticist movement is recognized by the use of pigmentary primary colors red, yellow, and blue in high saturation and combined with black and white in different spatial planes. The best examples of Neoplasticist architecture and design can be found in the works by Gerrit Rietveld (Fig. 3). The followers of the organicist architecture (whose paradigm is Frank Lloyd Wright), supporters of not covering surfaces with paints but leaving the materials to express their inherent color, were not less aware of the value of color. Perhaps with the aim of providing a rich polychromy without using paints, Wright made a plentiful use of stained glass windows, as a way of “painting” with light.

The main idea behind the polychromy of modern architecture is that color is not used with a decorative purpose. Instead, it should appear to enhance or modify the space, to express functions,



Polychromy, Fig. 3 G. Rietveld: Façade and interior of the Schröder house, Utrecht, The Netherlands, 1924

or to differentiate materials or construction elements. Within some architectural trends or contexts of the modern movement it was evident already in the early 1960s that color entailed new solutions but also new problems. It is for this reason that the Scientific and Technical Center for Building (Centre Scientifique et Technique du Batiment) in France organized a seminar on architectural polychromy under the direction of Guy Habasque, and with the participation of color consultants Antoine Fasani, Jacques Fillacier, Bernard Lassus, Georges Patrix, and Erasme Saffre. The discussions of this seminar were published in 1961 in an issue of the *Cahiers* of this Center [21].

After the modern movement (represented mainly by Le Corbusier, Frank Lloyd Wright, Mies van der Rohe, and Walter Gropius), the postmodernist reaction of the 1970s and 1980s was accompanied by a different view about color and polychromatic compositions. Postmodernism brought about a host of architects concerned with the references to history and the environment, and polychromy in architecture also acquired a new meaning under these orientations. As examples, the works by Charles Moore, Robert Venturi, Michael Graves, and Stanley Tigerman (in the USA), Paolo Portoghesi and Aldo Rossi (in Italy), Aldo van Eyck (in the Netherlands), and Mario Botta (in Switzerland) can be mentioned. In Latin America, there have been also outstanding cases of recognized masters of architecture who are paradigmatic for their conception and application of polychromy. Three of them deserve a special mention: Carlos Raúl Villanueva in Venezuela, Luis Barragán in Mexico, and Clorindo Testa in Argentina [5].

Cross-References

- [Color Circle](#)
- [Color Combination](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Complementary Colors](#)
- [Mosaics](#)
- [Pigments](#)
- [Primary Colors](#)

References

1. Kiilerich, B.: Monochromy, dichromy and polychromy in Byzantine art. In: Searby, D., et al. (eds.) ΔΩΡΟΝ ΡΟΔΟΠΟΙΚΙΑΟΝ. Studies in Honour of Jan Olof Rosenqvist, pp. 169–183. Uppsala University, Uppsala, Sweden (2012)
2. Nemcsics, A.: Colour Dynamics. Akadémiai Kiadó, Budapest (1990). (1993)
3. Goodwin, T.G.: A Short Account of the Art of Polychrome Historical and Practical. Joseph Masters, London (1860)
4. Hittorff, J.-L.: Restitution du temple d'Empédocle à Sélinonte, ou l'architecture polychrome chez les Grecs. Firmin Didot, Paris (1851)
5. Caivano, J.: Research on color in architecture and environmental design: Brief history, current developments, and possible future. Color. Res. Appl. **31**(4), 350–363 (2006)
6. Quatremère de Quincy, A.: Le Jupiter olympien; ou l'art de la sculpture antique . . . ouvrage qui comprend un essai sur le goût de la sculpture polychrome . . . et l'histoire de la statuaire en or et ivoire chez les grecs et les romains. Firmin Didot, Paris (1814)
7. Manzelli, V.: La policromia nella statuaria greca arcaica. L'Erma di Bretschneider, Rome (1994)
8. Kiilerich, B.: Towards a 'polychrome history' of Greek and Roman sculpture. J. Art Hist. **15**, 1–18 (2016)
9. Solon, L.V.: Polychromy: Architectural and Structural Theory and Practice. The Architectural Record, New York (1924)
10. Jones, O.: An Attempt to Define the Principles Which Should Regulate the Employment of Colour in the Decorative Arts. G. Barclay, London (1852)
11. Jones, O.: An Apology for the Colouring of the Greek Court in the Crystal Palace, with Arguments by G.H. Lewes and W. Watkiss Lloyd, and a Fragment on the Origin of Polychromy by Professor Semper. Crystal Palace Library and Bradbury & Evans, London (1854)
12. Jones, O.: The Grammar of Ornament. Day and Son, London (1856)
13. Audsley, W.J., Audsley, G.A.: Polychromatic Decoration as Applied to Buildings in the Medieval Styles. Henry Sotheran & Co., London (1882)
14. Ozenfant, A., Le Corbusier: Après le Cubisme. Edition des Commentaires, Paris (1918). Spanish translation, Después del Cubismo. In: Pizza, A. (ed.) Acerca del Purismo. Escritos 1918–1926, pp. 8–47. El Croquis Editorial, Madrid (1994)
15. Ozenfant, A.: Le Corbusier: Le Purisme. L'Esprit Nouveau 4 (1921). Nature et creation. L'Esprit Nouveau 19 (1923). Idées personnelles. L'Esprit Nouveau 27 (1924). Spanish translation. In: Pizza, A. (ed.) Acerca del Purismo. Escritos 1918–1926. El Croquis Editorial, Madrid (1994)

16. Le Corbusier: Polychromie architecturale. Foundation Le Corbusier, Paris, manuscript (1931). Posthumous edition. In: Rüegg, A. (ed.) Polychromie architecturale. Les claviers de couleurs de Le Corbusier de 1931 et de 1959. Birkhäuser, Basel (1997)
17. Le Corbusier: Des canons, des munitions? Merci! Des logis. ... S.V.P. Editions de L'Architecture d'Aujourd'hui, Paris (1938)
18. Cram, N.: It was never white, anyway: despite his reputation as the godfather of white architecture, Le Corbusier developed complex ideas about color. *Architecture (USA)*, pp. 88–91 (February 1999)
19. Schindler, V.M.: Prefabricated rolls of oil paint: Le Corbusier's 1931 colour keyboards. In: Caivano, J. (ed.) AIC 2004 Color and Paints. Proceedings of the Interim Meeting of the International Color Association, Porto Alegre, pp. 198–202 (2015). <https://aic-color.org/resources/Documents/aic2004.pdf>. Accessed 24 Mar 2020
20. De Heer, J.: The Architectonic Colour: Polychromy in the Purist Architecture of Le Corbusier. 010 Publishers, Rotterdam (2009)
21. CSTB: Centre Scientifique et Technique du Batiment: La Polychromie Architecturale. Cahier Nr. 423. CSTB, Paris (1961)

Precious Stones and Metals at Origins of Linguistic Expression of Color

► [Materials and the Origins of Abstract Color Naming](#)

Prelinguistic Color Categorization/Categorical Perception

► [Infant Color Categories](#)

Premetallized Acid Dyes

► [Dye, Metal Complex](#)

pRGCs

► [Melanopsin Retinal Ganglion Cells](#)

Primary Colors

Paul Green-Armytage
School of Media, Creative Arts and Social Enquiry, Curtin University, Perth, Australia

Synonyms

[Basic colors](#); [Elementary colors](#); [Fundamental colors](#); [Primitive colors](#); [Principal colors](#); [Pure colors](#); [Simple colors](#); [Unique hues](#)

Definition

There are two kinds of primary colors that are often confused or conflated. They can be called “physical primaries” and “visual primaries.” Sets of three “physical primaries” in the form of lights, paints, or inks can be mixed to produce a comprehensive range of other colors. These sets are “primary” because they have been found to deliver the most extensive and useful range of other colors. Colors are also recognized as “primary” by virtue of their appearance. These are the “visual primaries.”

Introduction

Much of the confusion that surrounds the topic of color can be exposed in a discussion about primary colors. Primary colors, for some people, are just colors that are particularly vivid. For those who work with color, there are two kinds of primary colors that are often confused or conflated. They can be called “physical primaries” and “visual primaries.”

Sets of three “physical primaries” in the form of lights, paints, or inks can be mixed to produce a

comprehensive range of other colors. These sets are “primary” because they have been found to deliver the most extensive and useful range of other colors. The particular lights or paints that work best can be recognized by their appearance: for lights, they appear orange-red, yellow-green, and violet-blue, but they are usually identified simply as red, green, and blue. The mixing process for lights is additive. For paints and inks, the mixing process is subtractive. Best results are obtained with paints or inks that appear purplish red, slightly greenish blue, and yellow. These are best identified as magenta, cyan, and yellow, but they are often loosely identified as red, blue, and yellow. The reds and blues for additive mixing appear different from those for subtractive mixing, but confusion can arise because of the range of different “reds” and “blues,” where an orangish red and a purplish red are both “red.”

Colors are also recognized as “primary” by virtue of their appearance. These are the “visual primaries.” In the continuous circle of hues, there are four that are recognized as “unique”: a yellow that appears neither greenish nor reddish, a red that appears neither yellowish nor bluish, a blue that appears neither reddish nor greenish, and a green that appears neither bluish nor yellowish. These visual primaries serve as conceptual reference points and, together with white and black, anchor the Natural Color System (NCS), which orders colors according to how they appear. Confusion can arise when visual primaries are thought of as the same thing as physical primaries. Confusion can be overcome when the distinction is maintained between visual appearances, on the one hand, and the lights, paints, and inks which give rise to those appearances, on the other hand.

Before going deeper into the two main questions – what are the primary colors and what makes them “primary” – it will be helpful to consider the nature of color itself.

The Nature of Color

Color means different things to different people. There are the colors that physicists measure, the colors that chemists analyze, the colors that

psychologists study, the colors that artists brush onto paper or canvas, and the colors that people talk about in general conversation. All these may be related, but they are not the same kind of thing.

It is possible to identify seven different “kinds of color” [1], but for present purposes, a broad distinction can be made between the physical aspects of color and color as it is experienced in perception. In general conversation, a rose or the setting sun might be described as red, and for most people that red would be a physical property of the rose or the sunlight. They might also recognize as red the subjective visual experience they have when they look at the rose or the setting sun. The one word “red” is used for two different kinds of thing. The lack of distinction between the physical and visual aspects of color is responsible for much of the confusion that exists in connection with color in general and with primary colors in particular.

Confusion Between the Physical and Visual Aspects of Color

The confusion is evident in what might be assumed to be an authoritative definition. In the *Concise Oxford Dictionary* [2], a primary color is defined as “any of the colours red, green, and blue, or (for pigments) red, blue, and yellow, from which all other colours can be obtained by mixing.” In this definition, the meaning of the word “color” changes halfway through. Pigments are physical substances. A particular pigment might appear red, but that redness is not the same thing as the pigment itself. It can help to remove the confusion if physical and visual are added after the word “color” to clarify the distinction between color meaning something physical, on the one hand, and color meaning visual experience, on the other hand. Then that definition could be reworded as follows: “any of the colors (physical) . . . from which all other colors (visual) can be obtained by mixing.”

The same change of meaning can be recognized in another definition, this from *The Color Compendium* [3]: “Simply stated, primary colors are ones that cannot be mixed or formed by any

combination of other colors.” This could also be reworded as follows: “. . . primary colors (visual) are ones that cannot be mixed or formed from any combination of other colors (physical).”

Evolution of Ideas About Primary Colors

Although the definitions above treat the physical and the visual aspects of color as if they were the same thing, they do illustrate the two main ideas about what should be understood by primary colors. One is that primary colors are physical and are used in mixing processes; the other is that primary colors are visual and serve as perceptual and cognitive reference points. The evolution of these ideas can be traced through history.

Greek Philosophers

In the fifth century BC, Empedocles recognized four principal colors as corresponding with the four elements: black, white, red, and green with earth, air, fire, and water [4, p. 15]. For Aristotle, white and black, combined in different proportions, were the sources of other colors. In a line between white and black, he identified five other principal colors – yellow, red, violet, green, and blue. “. . . from these all others are derived by mixture.” [5, p. 698].

Leonardo da Vinci

Closer to the present time, but with echoes of Empedocles, Leonardo da Vinci wrote that

white is the first among the simple colors, and yellow the second, green the third of them, blue is the fourth, and red is the fifth, and black the sixth. And white is given by light, without which no color may be seen, yellow by earth, green by water, blue by air and red by fire, and black by darkness . . . [6, p. 70].

These six colors would seem to correspond with visual primaries as reference points. On the mixture of colors Leonardo wrote,

I call simple colours those which are not compounded and cannot be compounded by means of a mixture of other colours . . . After black and white come blue and yellow, then green and tan, that is to say, tawny, or if you wish to say

ochre, and then deep purple and red. These are the eight colours and there are no more in nature, and with these I begin the process of mixing, first with black and white, and then black with yellow . . . [6, p. 72]

It seems likely that these eight colors represent the pigments which Leonardo used in his paintings. His eight simple colors would correspond with physical primaries, but their number would have had more to do with the limitations of available pigments and his own approach to the practicalities of painting than any of the theories of primary colors that are more familiar today.

Robert Boyle

Something much closer to present ideas about mixing a comprehensive range of colors (visual) from a limited set of colors (physical) can be found in the writings of Robert Boyle:

. . . there are but few Simple and Primary Colours (if I may so call them) from whose Various Compositions all the rest do as it were Result. For though Painters can imitate the Hues (though not always the Splendor) of those almost Numberless differing Colours that are to be met with in the Works of Nature, and of Arte, I have not yet found, that to exhibit this strange Variety they need employ any more than *White*, and *Black*, and *Red*, and *Blew*, and *Yellow*; these *five*, Variously *Compounded*, and (if I may so speak) *Decomposed*, being sufficient to exhibit a Variety and Number of Colours, such, as those that are altogether Strangers to the Painters Pallets, can hardly imagine. [7, pp. 219–220]

There are two significant features of this passage. First, it shows that the possibility of producing a complete sequence of hues with only three paints (physical colors) was well established. Second, it introduces, by implication, the concept of color gamut. A gamut is the range of colors (visual) that can be produced with a given set of colors (physical). Although with Boyle's *Red*, *Blew*, and *Yellow* it was possible to produce all the hues, it was not possible to produce all the colors – colors of greater “splendor” were out of reach. With *Red* and *Blew*, it was possible to produce a range of purples but not the vivid purples of some flowers. The word “hue” is often used as a synonym for “color,” but in a discussion of primary colors, it is important to recognize that

hue is just one of the dimensions of color. Boyle's "splendor" is another dimension which refers to how vivid a color may be. Words used today for this dimension include intensity and saturation as well as the more specialist terms "chroma" and "chromaticness."

Jakob Christoph Le Blon

While Boyle was a scientist, Jakob Christoph Le Blon was an artist. Le Blon developed a system of printing with transparent inks from three or four plates on white paper that is essentially the same as the process used by printers today. Three plates, one inked with yellow, one with red, and one with blue, were printed one over the other. "Painting can represent all *visible* Objects, with three Colours, *Yellow, Red, and Blue*; for all other Colours can be compos'd of these *Three*, which I shall call *Primitive*" [8]. Sometimes Le Blon used a fourth plate, inked with black, much as printers today use black to deepen the colors and sharpen the image.

Isaac Newton

Isaac Newton experimented with prisms and demonstrated that different wavelengths of light are refracted by a prism to a greater or lesser extent and that the refracted light appears in different colors. When projected onto a white surface, the refracted light is spread out to reveal the colors of the spectrum. In Newton's words, "The Light whose Rays are all alike Refrangible, I call Simple, Homogeneal and Similar . . . The Colours of Homogeneal Lights, I call Primary, Homogeneal and Simple . . ." [9, p. 4]. For Newton, therefore, a primary color was light of a single wavelength.

The spectrum can be divided and subdivided many times over, so the number of primary colors in this physical sense is very large indeed. But Newton chose to name just seven: "The original or primary colours are red, yellow, green, blue and a violet-purple together with orange, indigo and an indefinite variety of intermediate gradations" (Newton, quoted in [10, p. 225]). Newton developed a circular diagram, divided into seven segments, which could be used to determine the likely appearance of a given mixture of his primary lights.

Thomas Young

The fact that it was possible to mix a comprehensive range of colors (visual) from a limited set of primary colors (physical) led George Palmer and Thomas Young, working independently, to similar insights about what there must be in the eye that enables humans to experience such a wide range of color appearances. Young's speculation has been confirmed in its essential details:

Now, as it is almost impossible to conceive each sensitive point of the retina to contain an infinite number of particles, each capable of vibrating in perfect unison with every possible undulation, it becomes necessary to suppose the number to be limited, for instance, to the three principal colors, red, yellow, and blue . . . and that each of the particles is capable of being put in motion less or more forcibly by undulations differing less or more from perfect unison . . . each sensitive filament of the nerve may consist of three portions, one for each principal color. [11, p. 147]

James Clerk Maxwell

Where Le Blon had demonstrated that an image could be reproduced in a print by the subtractive mixing of three transparent inks, James Clerk Maxwell demonstrated how an image could be reproduced by the additive mixing of colored light. The first color photograph, taken by Thomas Sutton, was used as an illustration for a lecture by Maxwell in 1861 [12, p. 321]. A ribbon had been photographed successively through a red, a green, and a blue filter. The three photographs were then projected through corresponding filters so that the three images were superimposed on the screen.

Maxwell also developed a means of measuring color. When a disk with segments in different colors is spun at high speed, the colors blend and a new color is seen. Three paper disks, one in each of the additive primary colors (physical), are interleaved in such a way that the visible proportions of each can be adjusted. A disk in the color that is to be measured is fixed over the three interleaved disks. The superimposed disks are then spun in a series of trials, with the proportions of the three primary colored disks being adjusted after each trial, until the combination of three primary colors matches the color to be measured. The measurement is then expressed in terms of the relative

amounts of each of the primary colors needed to achieve the match. In this process, colors (physical) and colors (visual) come together. The colored papers are physical, but it requires a visual judgment to establish a match.

Ewald Hering

During the nineteenth century, the idea that color vision depends on three different kinds of sensitive cell in the eye, with three corresponding primary colors, became established. It was challenged by Ewald Hering, who used his own observations as a starting point. He concluded that there are six fundamental colors (Urfarben) that must have their counterpart in the nervous system associated with vision. Hering's Urfarben can be called primary colors (visual). The claim is that any color can be described in terms of its relative resemblance to these Urfarben which serve as reference points. They are the basis of the Swedish standard Natural Color System [13]. After black and white, Hering defines what are now recognized as the "unique hues":

... there are four outstanding loci in the series of hues that make up the closed circle: first the locus of the yellow that shows no remaining trace of redness, and yet reveals no trace of green; second the locus of the blue for which the same is true. These two hues may be called *primary yellow* and *primary blue* (Urgelb and Urrblau). Likewise we can name, third, the red, and, fourth, the green that are neither bluish nor yellowish primary red and primary green (Urrot and Urrgrün). [14, p. 42]

The CIE System

In 1931, the International Commission on Illumination (CIE) established the system of color measurement that is today's world standard [15, pp. 59–63]. All possible colors (physical) can be plotted on the CIE chromaticity diagram. The position of a color on the diagram is determined by the relative output of three primary lights that, when mixed together, would match the color. The measurement can be expressed in terms of tristimulus values. As with Maxwell's system of spinning disks, the matches that underpin the CIE system were initially the result of visual judgments made by human observers. But there is no set of three colored lights (physical) that can

match all possible colors (visual). The most vivid colors (visual) are those of Newton's homogeneous lights – the pure colors of the spectrum. In order to measure such colors, it was necessary to reduce their "splendor" by introducing negative values. This led to complexities that were overcome by clever mathematics and the introduction of theoretical "super light sources" that could not be physically realized but that were able to deliver matches for the spectral colors.

Primary Colors in the Twenty-First Century

The different ways in which the term "primary colors" is used today can be related to these ideas that evolved in the past.

Further Problems with Standard Definitions

The definitions of primary colors, quoted in the section "[Confusion Between the Physical and Visual Aspects of Color](#)" above, not only confuse the physical and visual aspects of color, but they are also misleading in other ways. As Boyle observed, and the CIE confirmed, there is no set of primary colors (physical) from which all other colors can be obtained by mixing. It is possible to obtain all the hues but not all the colors – only those within a limited gamut. The CIE had to resort to its theoretical super lights. A full gamut of all other colors, right out to Newton's homogeneous lights, can only be achieved in theory, with the CIE super lights, and not in practice.

The second definition, that primary colors cannot be mixed from a combination of other colors, is also misleading. It would be true of the CIE theoretical super lights but not in the case of normal lights, paints, or inks. It would make no sense to claim that primary colors (physical) cannot be mixed from a combination of other colors (physical). The primary colors that do make sense in the context of this definition are visual colors. Leaving aside black and white, the visual primaries should be the unique hues – Hering's *Urfarben* – and there is nothing unmixable about these. A paint that appears yellow can be mixed with paint that appears magenta to produce a red

that is neither bluish nor yellowish and with a paint that appears blue to produce a green that is neither bluish nor yellowish.

Primaries for Mixing Other Colors in Print and on Screen

Primary colors (physical) for mixing a large range of other colors (visual) have particular application for the reproduction of colored images in print and on the television or computer screen. There is scope for further confusion here because the mixing processes are not the same. When layers of transparent ink are superimposed in a print, the result is a subtractive mixture. Additive mixture occurs when pools of colored light are projected to overlap or when the small points of colored light blend on a screen.

It was not until the nineteenth century that the difference between subtractive and additive processes was recognized. Then it became clear that the primaries (physical) for each process were not only different in that pigments and dyes are physically different from colored lights but also that the most useful sets of three colors (physical) appear different. The two sets are listed in the first definition set out in the section “[Confusion Between the Physical and Visual Aspects of Color](#).” Additive primaries are listed first, then the subtractive primaries. They are described as red, green, and blue or (for pigments) red, blue, and yellow. This suggests that the only difference between the two sets is that one set has green and the other set has yellow. However, the lights that work best for additive mixture appear orange-red, yellow-green, and violet-blue, while the inks for subtractive mixture appear purplish red (magenta), slightly greenish blue (cyan), and yellow. The differences between the orange-red and magenta, and between the violet-blue and cyan, are masked by the way the names “red” and “blue” stretch to cover a range of different hues. What makes these two sets primary is that they have been found to deliver the most useful gamuts of other colors (visual). The gamuts could be increased for each process by adding further inks or lights, but any slight improvement would not justify the very large extra costs that this would involve.

Mixing with a Larger Number of Colors (Physical)

When a color is required in a print that is beyond the reach of the cyan, magenta, and yellow process inks, it may be achievable with a mixture of inks formulated according to the Pantone Matching System[®]. This system has 14 separate inks that can be premixed before being used on the press.

A similar number of different colors (physical) are used to mix the very large range of different colors (visual) that are available for painting the interior and exterior of buildings. The tinting machines for Dulux Australia have four main bases and 13 different tints that add pigments to a base according to given formulas.

Today’s artists, like Leonardo before them, need not restrict themselves to a set of three primary colors (physical). They use as many different paints as they need to achieve the results they want.

Primary Confusion

The confusion surrounding the different notions of primary colors, especially that between primary colors (physical) and primary colors (visual), was the subject of a paper that Green-Armytage gave at a conference on color education [16]. This confusion is especially evident in what he calls traditional art school color theory.

Among the most widely used and influential books on color theory for artists are *The Art of Color* [17] and the condensed version *The Elements of Color* by Johannes Itten [18]. Itten’s account should be read carefully [17, p. 34, 18, p. 29]:

... let us develop the 12-hue color circle from the primaries – yellow, red and blue. As we know, a person with normal vision can identify a red that is neither bluish, nor yellowish; a yellow that is neither greenish nor reddish; and a blue that is neither greenish, nor reddish.

These definitions are reminiscent of Hering’s definitions except that green is not included as a primary. However, green is included in the definitions of primary yellow and primary blue. Itten did not (could not?) describe primary blue as

neither reddish nor *yellowish* or primary yellow as neither reddish nor *bluish*.

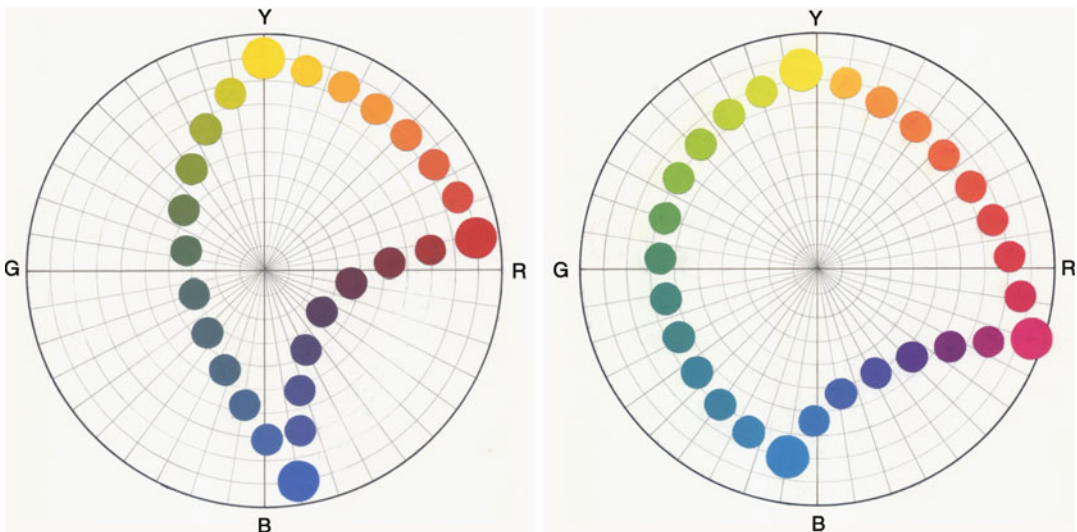
Itten goes on to describe secondary colors as "... three mixed colors, each composed of two primaries." Later in the books [17, p. 136, 18, p. 88], he describes green as "... the intermediate between yellow and blue. According as green contains more yellow or more blue, the character of its expression changes." Did he mean that there could be more or less yellow or blue paint in the mixture which resulted in the perception of green? Or did he mean that in the green percept there might be more or less resemblance to yellow and blue as primary colors (visual)?

It is not clear from the books whether Itten associated his primary yellow, red, and blue with any particular pigments. There is certainly no claim that with three such paints all other colors can be obtained by mixing. But what is clear from experiment is that three paints which might fit Itten's definitions would not work so well for producing a comprehensive range of other colors (visual) as would paints that look the same or similar to the cyan, magenta, and yellow process inks used by printers. Itten's primaries would not be a good choice as primary colors (physical).

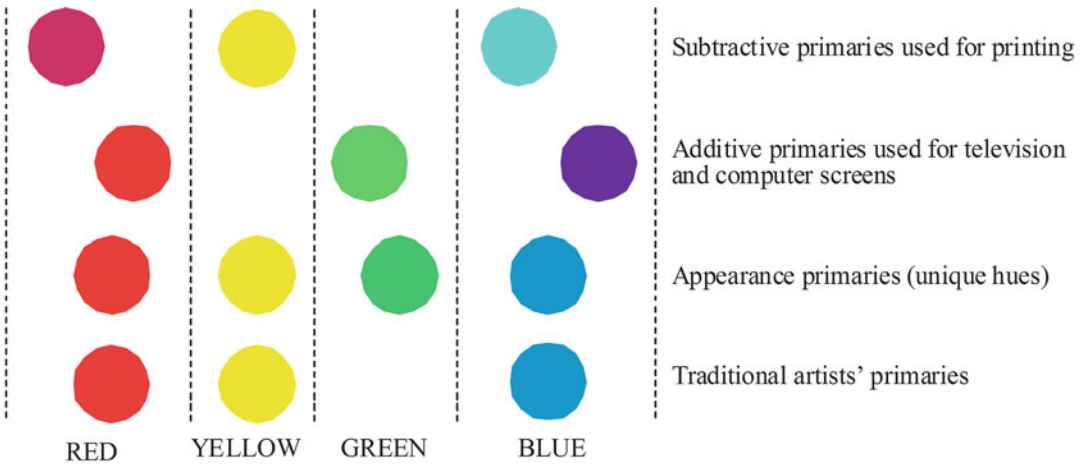
And without green they are not satisfactory as primary colors (visual).

Plotting the Gamut for a Given Set of Colors (Physical)

The color circle of the Natural Color System (NCS) can be used as a diagram on which the appearance of paints can be plotted. The unique hues, as reference points, are spaced at equal intervals round the circle. The achromatic colors – the pure whites, grays, and blacks – are located in the center of the circle. The most vivid colors (visual) that can be imagined are located on the circumference. The less vivid colors (visual) are located between the circumference and the center, closer to the circumference or to the center as they are more or less vivid. Points on the diagram representing two paints can be connected by mixture lines to indicate the changing appearance of the mixtures as they contain more or less of each paint. The mixture lines connecting three paints trace an irregular triangle and mark the gamut of appearances that can be achieved with that set of paints. This makes it possible to compare the relative merits of different sets of primary colors (physical). The gamuts achievable with different sets of "primaries" are illustrated in Fig. 1.



Primary Colors, Fig. 1 Mixture lines connecting two sets of "primary" paints plotted on the *color circle* of the Natural Color System (NCS)



Primary Colors, Fig. 2 Four sets of primary colors identified loosely as “red,” “yellow,” “green,” and “blue” to show how use of color names contributes to the way the sets are confused

On the left, cadmium yellow pale, cadmium red, and ultramarine are typical of the paints regarded as “primary” in traditional art school color theory. While these paints can deliver all the hues, the greens and purples are very dull. If more vivid greens and purples are required, an artist would need to include more paints on the palette. A more extensive gamut can be achieved with the addition of a turquoise or cerulean blue and a vivid magenta such as Bengal rose. The paints used on the right were lemon yellow, Bengal rose, and turquoise blue. These paints appear similar to the yellow, magenta, and cyan process inks as used in printing. Their status as preferred primaries for subtractive mixing is clear from the more extensive color gamut that they can deliver. Within the mixture lines are the colors (visual) that could be achieved with that set of paints. If a desired color (visual) would be located outside the mixture lines, there may be a paint that could be added that would bring that color (visual) within reach [19].

The Role of Color Names in the Confusion of Different Kinds of Primaries

In the section “[Primaries for Mixing Other Colors in Print and on Screen](#)” above, it is argued that the difference between primaries for additive and subtractive mixing is masked by the way that the names “red” and “blue” are each used to describe

a range of different hues. The only point of difference between the two sets of primaries seems to be that green is a primary for additive mixing and yellow for subtractive mixing; both sets include “red” and “blue.” It might seem that the unique hues, as visual primaries, resolve the conflict by including both yellow and green, but this only adds to the confusion. And further confusion comes from the common expectation of art students that they can “mix all the other colors” from the primaries red, yellow, and blue. The red, yellow, and blue of traditional art school color theory are typically defined in terms used for the visual primaries, the unique hues. The four sets of primaries are compared in Fig. 2.

Conclusion

There has been a long history of singling out particular colors as being somehow special. The realization that a comprehensive range of colors (visual) can be mixed from a limited set of paints, inks, dyes, or lights led to the development of ways to reproduce fully colored images in prints and projected photographs, techniques pioneered by Le Blon and Maxwell. This realization also inspired ideas about the physiology of the eye and how it is that humans can experience such a great variety of colors.

Concepts of primary colors can help people to work successfully with colors. These concepts can also be confused. The key is to recognize that there are two kinds of primary colors: physical primaries, which take the form of lights, paints, inks, and dyes to be used in mixing a comprehensive range of other colors (visual), and visual primaries as very specific visual experiences – pure white, black, and the four unique hues. By using the visual primaries for reference, the physical processes involved in the production and reproduction of colors, which may involve the physical primaries, can be monitored and controlled.

Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [Color Circle](#)
- [Dye](#)
- [Newton, \(Sir\) Isaac](#)
- [Psychologically Pure Colors](#)
- [Unique Hues](#)

References

1. Green-Armytage, P.: Seven kinds of colour. In: Porter, T., Mikellides, B. (eds.) *Colour for Architecture Today*. Taylor and Francis, Abingdon (2009)
2. Thompson, D.E.: *The Concise Oxford Dictionary*. Oxford University Press, Oxford (1995)
3. Hope, A., Walch, M.: *The Color Compendium*. Van Nostrand Reinhold, New York (1990)
4. Beare, J.: *Greek Theories of Elementary Cognition*. The Clarendon Press, Oxford (1906)
5. Aristotle: *Sense and Sensibilia*. In: Barnes, J. (ed.) *The Complete Works of Aristotle*. Princeton University Press, Princeton (1984). [4th c. BC].
6. da Vinci, L.: *Leonardo on Painting*. Yale University Press, New Haven (1989 [1518]).
7. Boyle, R.: *Experiments and Considerations Touching Colours*. Johnson Reprint Corporation, New York (1964 [1664]).
8. Le Blon, J.C.: *Coloritto*. Van Nostrand Reinhold, New York (1980 [1756])
9. Newton, I.: *Opticks*, 3rd edn. Bell & Sons, London (1931 [1730]).
10. McLaren, K.: Newton's indigo. *Color Res. Appl.* **10**(4), 225–229 (1985)
11. Young, T.: On the theory of light and colours. In: Peacock, G. (ed.) *Miscellaneous Works of the Late Thomas Young*. John Murray, London (1972 [1801])
12. Kemp, M.: *The Science of Art*. Yale University Press, New Haven (1990)
13. Hård, A., Sivik, L.: NCS – Natural Color System: A Swedish standard for color notation. *Color Res. Appl.* **6**(3), 129–138 (1981)
14. Hering, E.: *Outlines of a Theory of the Light Sense*, trans. Hurvich, L.M., Jameson, D. Harvard University Press, Cambridge, MA (1964 [1905])
15. Berns, R.: *Billmeyer and Saltzman's Principles of Color Technology*, 3rd edn. Wiley, New York (2000)
16. Green-Armytage, P.: Primary confusion. In: Arnkil, H., Hämäläinen, E. (eds.) *Aspects of Colour*. The University of Art and Design Helsinki UIAH, Helsinki (1995)
17. Itten, J.: *The Art of Color*. Van Nostrand Reinhold, New York (1961)
18. Itten, J.: *The Elements of Color*. Van Nostrand Reinhold, New York (1970)
19. Green-Armytage, P.: Paints as 'magnets' to guide the mixing process. In: Hansuebsai, A. (ed.) *Proceedings AIC 2003 Bangkok Color Communication and Management*, pp. 309–311. Chulalongkorn University, Bangkok (2003)

Primate Color Vision

- [Evolution of Primate Color Vision](#)

Primitive Colors

- [Primary Colors](#)

Principal Colors

- [Primary Colors](#)

Printing

- [Coloration, Textile](#)

Problems of the Earliest Language Development in the Domain of Color

► Evolution of Color Expression

Protanopia

Galina V. Paramei¹ and David L. Bimler²

¹Department of Psychology, Liverpool Hope University, Liverpool, UK

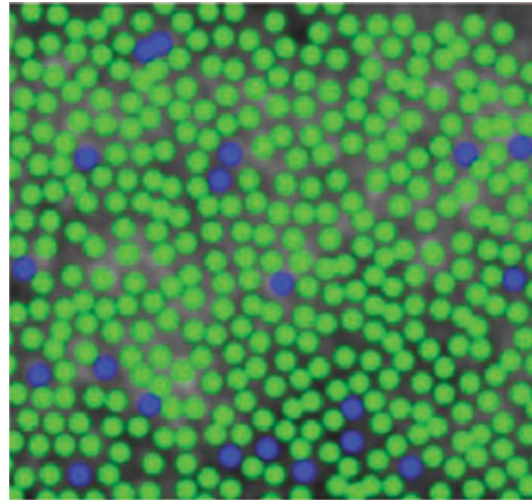
²School of Psychology, Massey University, Palmerston North, New Zealand

Synonyms

Color blindness; Congenital color vision abnormality; Daltonism; Dichromacy; “Red-blindness”; Red-green deficiency; X-linked inherited color vision deficiency

Definition

Protanopia (from the Greek *protos*: first + *an*: not + *opia*: a visual condition) is a congenital form of severe color deficiency (dichromacy) affecting the red-green opponent color system. Protanopes do not distinguish between colors along a specific direction (protanopic confusion lines) in color space and are able to match all colors using two primaries (unlike trichromats who require three primaries). Protanopia is linked to the X-chromosome and arises from loss or alteration of the gene encoding the opsin of the long-wavelength (L-) photopigment; this malstructure results in the absence of functioning L-cones in the retina. Protanopia follows a recessive pedigree pattern, with incidence of ca. 1% in the Caucasian male population. It is distinct from protanomaly in which some degree of discrimination among such colors remains.



Protanopia, Fig. 1 Adaptive optics retinal imaging: Pseudocolor image of the protanopic cone mosaic; all cones are of the M- (“green”) or S- (“blue”) type. (Source: Carroll et al. [2]; Fig. 4b, p. 8465. Copyright (2004) National Academy of Sciences, USA)

Retinal Mosaic

At the level of retinal function, protanopia is the absence of L-cones (Fig. 1) (► [Cone Fundamentals](#); ► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)), i.e., cone cells expressing a photopigment containing an L-opsin component, tuned for selective sensitivity to longer wavelengths such as orange and red light [1].

Genetic Background

Functional loss of the L-cone photoreceptor cells is caused by a mutant cone photopigment gene on the X-chromosome (► [Genetics of Photoreceptors, Genetics and Color Vision Deficiencies, Genes of Cone Photopigments, Genes and Cones](#)). The gene coding for the L-opsin may be missing altogether from a protanope’s genome; it may be replaced by a second copy of the M-opsin (sensitive to slightly shorter wavelengths), or it may be present but damaged so that the protein is not expressed or cannot be transported to the ciliary segment of cone cells expressing that gene [2–4].

When present, the L-opsin gene is located on the X-chromosome, followed by one or more copies of the M-opsin gene (► [Genetics and Color Vision Deficiencies, Genes of Cone Photopigments, Genes and Cones](#)). This architecture, and the high similarity of the two genes, facilitates mismatching of DNA during meiotic crossover, leading to deleterious mutations of the L-opsin encoding gene [3, 4].

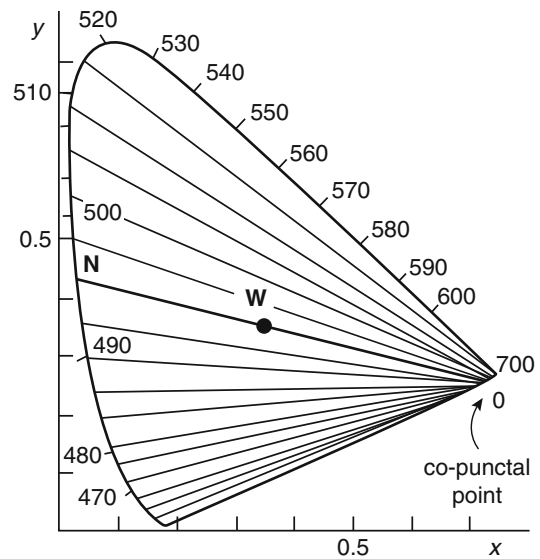
With two X-chromosomes, females are far less prone to protanopia. Notably, mothers and daughters of protanopes can be heterozygous for protanopia (i.e., carriers of the condition) with one normal and one aberrant X-chromosome; the former is active in a portion of their cone cells, providing enough L-cones to sustain normal trichromacy. However, minor departures from normality are discernible [5], including reduced sensitivity to long wavelengths, known as Schmidt's sign, and reduced subjective prominence of red-green dissimilarity [6].

Confusion Lines

L-signal/M-signal comparison in bipolar ganglion color-comparison cells is not present in protanopes (Color Vision, Opponent Theory); hence there is no difference in appearance of any colors which differ only in the L–M signal. That is, although protanopes lack the L- or “red” cones, they are not “red blind” in a literal sense but rather are red-green blind. Examples of colors confused by protanopes are red, yellow and green or cyan, gray, and purple. Plotted in a chromaticity diagram (► [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength; ► CIE Chromaticity Coordinates \(xyY\)](#)), the confused colors lie on straight “confusion lines” that converge at the protanope copunctal point (Fig. 2).

Luminous Efficiency, Luminance Effect, and Wavelength Discrimination

In the absence of L-cones tuned to longer wavelengths, protanopes display a distinctive insensitivity to 550–750 nm lights (orange and red)



Protanopia, Fig. 2 Confusion lines indicate colors that cannot be distinguished by a protanope. The line going through the white point (*W*) indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing L-cone fundamental. Drawn in the CIE 1931 chromaticity diagram [7]

[8]. Their luminosity function has peak sensitivity around 540 nm compared to the maximum near 555 nm of normal trichromats, i.e., it is blue-shifted toward the $\lambda_{\max}$ of the M-cone pigment [1, 9]. Phenomenally, reds appear relatively dark for protanopes and are confused with grays and bluish turquoise.

Thus, protanopes attend more to luminance changes than normal trichromats as a color cue, compensating for their poor hue discrimination [10]. Notably, at lower photopic and mesopic luminance levels, protanopes' color discrimination slightly improves, perhaps reflecting input from rod cells [11].

Hue discrimination, as a function of wavelength, is considerably worse in protanopes than in normal trichromats. Their lowest discrimination thresholds ($\Delta\lambda$), approaching the trichromatic value, are around the protanopic “neutral point,” ca. 490 nm (see below). $\Delta\lambda$ rises on both sides of this minimum, increasing rapidly on the short-wavelength side and on the long-wavelength side until wavelengths are not distinguished beyond approximately 550 nm [8].

Color Perception

A protanope recognizes far fewer distinct colors than a trichromat (► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)). Instead of all possible equiluminant colors being arranged in a two-dimensional plane (► [Psychological Color Space and Color Terms](#)), for a protanope they form a single-dimensional line, running between the two extremes of saturated blue and saturated yellow, separated by a “neutral zone” of subjective gray, with midpoint at ca. 490 nm of spectral light (blue-green for normal trichromats) [12]. Protanopes distinguish approximately 21 steps along this one-dimensional line, each corresponding to a separate band of interchangeable colors [1]. The usually separate chromatic qualities of “saturation” and “hue” are thus combined along this line. This protanopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

Whether the subjective qualities of the two extremes are identical to the normal trichromatic “blue” and “yellow” remains controversial and rests on reports from a handful of individuals with unilateral dichromacy. The blue-yellow convention is used in dichromacy simulation software, which converts color scenes to show how a protanope would see them [13].

Despite this model of a two-hue palette for self-luminous colors, the color percepts of protanopes remain in question. Their object-color palette was proved to include all six Hering component colors (red, green, blue, yellow, white, and black) [14]. However, in contrast to normal trichromats, protanopes perceive only a weak green component in object-colors and do not experience “unique green,” while their red and green components are confounded with an achromatic element (► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)).

Diagnosis of Protanopia

A number of tests for protanopia are available, varying in sensitivity, selectivity, and ease of use

[15, 16] (► [Color Vision Testing](#)). Pseudoisochromatic plates and arrangement/panel tests, widely used in survey/field studies, directly address a protanope’s confusions between color pairs along the confusion axis.

In the pseudoisochromatic tests (Ishihara plates, Hardy-Rand-Rittler test), each diagnostic plate shows a normal observer design (► [Pseudoisochromatic Plates](#); ► [Color Vision Testing](#)) – one- or two-digit numbers or curved lines defined by a color difference – which disappears for protanopes, so that the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction.

In the arrangement tests (► [Color Vision Testing](#)) (Farnsworth-Munsell 100 Hue test, Farnsworth D-15 test), the one-dimensionality of a protanope’s color “plane” collapses into a line (► [Color Circle](#)), causing drastic departures from a normal observer’s sequence. Errors, or cap transpositions “distances,” are summed to quantify the severity of protanopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the protanopic angle of the confusion axis (► [Color Perception and Environmentally Based Impairments](#)).

The “gold standard” of diagnosis of protanopia is the Rayleigh match on a small viewing field (2°) of the Nagel anomaloscope (► [Color Vision Testing](#)) [16]. Unlike normal trichromats who reproducibly choose a unique ratio of red and green to match the yellow intensity, a protanope can match *any* red/green ratio to the yellow light by adjusting the luminance of the latter and dimming it if the mixed light is mainly red, while a green-dominated mixture requires for a protanope a more luminant yellow. Notably, some protanopes, when setting Rayleigh matches tested with a large viewing field (10°), perform as anomalous trichromats. Various factors related to retinal eccentricity are presumed, e.g., rod intrusion, S-cone intrusion, spatial variation in the macular pigment, etc. [1].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour Test [17] and the Colour Assessment and Diagnosis test [18] (► [Color Vision Testing](#)), interactively quantify any loss of chromatic sensitivity

along the protanopic confusion line. For a protanope, in the frame of the CIE ($u'v'$) 1976 chromaticity diagram (► [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the protanope copunctal point [17].

Incidence of Protanopia

The prevalence of inherited protanopia varies between human populations [19]. For Caucasian populations, the reported incidence of protanopia is 1.01% in males and 0.002% in females [1]. Recent surveys of red-green deficiency in populations of different racial origin found the incidence of protanopia in Caucasian males to vary between 0.96% and 1.27%. In a marked difference, the incidence is about half this rate in indigenous populations in Asia, Africa, Australia, and Central America but is observed to rise in geographic areas that have been settled by incoming migrants, in particular in coastal areas visited by Europeans in historical times [19]. The increasing prevalence of congenital red-green deficiencies, including protanopia, has tentatively been explained by a relaxation in selection pressures as hunting and gathering cultures evolve toward industrialized societies [1]. More recently, however, it is rather attributed to genetic drift and founder events (variable sampling of the gene pool contingent on the population size) [19].

Cross-References

- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision Testing](#)
- [Cone Fundamentals](#)
- [Cone Pigments](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)

References

1. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Opsin genes, cone pigments, color vision and colorblindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: from Genes to Perception*, Ch. 1, pp. 3–52. Cambridge University Press, Cambridge (1999)
2. Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D. R.: Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8461–8466 (2004)
3. Neitz, M., Neitz, M.: The genetics of normal and defective color vision. *Vis. Res.* **51**, 633–651 (2011)
4. Hunt, D.M., Carvalho, L.S.: The genetics of color vision and congenial color deficiencies. In: Kremers, J., Baraas, R.C., Marshall, N.J. (eds.) *Human Color Vision*, Ch. 1, pp. 1–32. Springer, Cham (2016)
5. Jordan, G., Mollon, J.D.: A study of women heterozygous for colour deficiencies. *Vis. Res.* **33**, 1495–1508 (1993)
6. Bimler, D., Kirkland, J.: Colour-space distortion in women who are heterozygous for colour deficiency. *Vis. Res.* **49**, 536–543 (2009)
7. Pitt, F.H.G.: *Characteristics of Dichromatic Vision* Medical Research Council Special Report Series, No. 200. His Majesty's Stationery Office, London (1935)
8. Hecht, S., Shlaer, S.: The color vision of dichromats. I. Wavelength discrimination, brightness distribution, and color mixture. *J. Gen. Physiol.* **20**, 57–82 (1936)
9. Pitt, F.H.G.: The nature of normal trichromatic and dichromatic vision. *Proc. R. Soc. Lond. B.* **132**, 101–117 (1944)
10. Paramei, G.V., Bimler, D.L., Cavonius, C.R.: Effects of luminance on color perception of protanopes. *Vis. Res.* **38**, 3397–3401 (1998)
11. Bimler, D.L., Paramei, G.V.: Bezold-Brücke effect in normal trichromats and protanopes. *J. Opt. Soc. Am. A.* **22**, 2120–2136 (2005)
12. Massof, R.W., Bailey, J.E.: Achromatic points in protanopes and deutanopes. *Vis. Res.* **16**, 53–58 (1976)

13. Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J.D.: What do colorblind people see? *Nature*. **376**, 127–128 (1995)
14. Logvinenko, A.: On the colours dichromats see. *Col. Res. Appl.* **39**, 112–124 (2014)
15. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
16. Pokorny, J., Smith, V.C., Verriest, G.: Congenital color defects. In: Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L.G. (eds.) *Congenital and Acquired Color Vision Defects*, pp. 183–241. Grune and Stratton, New York (1979)
17. Regan, B., Reffin, J.P., Mollon, J.D.: Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vis. Res.* **34**, 1279–1299 (1994)
18. Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour vision: consequences for occupational environments. In: Best, J. (ed.) *Colour Design: Theories and Application*, pp. 24–82. Woodhead Publishing Ltd., Cambridge (2012)
19. Birch, J.: Worldwide prevalence of red-green color deficiency. *J. Opt. Soc. Am. A*. **29**, 313–320 (2012)

Pseudoisochromatic Plates

Marisa Rodriguez-Carmona
Optometry and Visual Science, City University
London, London, UK

Synonyms

Color vision test; Isochromatic color confusions; Screening plates

Definition

Printed pseudoisochromatic plates are the most widely used type of color vision test to screen for color vision deficiency. The principle is that the color of a target (digit or letter) embedded in a background of another color appears “falsely of the same color” to color-deficient people. Target and background chromacities are chosen carefully to be the ones confused by people with color vision deficiency.

Design of Pseudoisochromatic Plates

Pseudoisochromatic plates must be carefully designed in order to be effective. As a result of the printing process, minor misalignments between a figure of one chromaticity and the uniform background of another undistinguishable chromaticity may reveal the shape of the digits or letters. In addition, the relative luminous efficiency of the eye varies considerably within normal trichromats and even more so in color-deficient observers. Hence, a figure and background that are equally light for one color defective are not necessarily so for another. The first pseudoisochromatic plates, introduced in Germany by Stilling in 1877, addressed these issues by breaking up the test target and background into a number of discrete patches each with its own shape and contour, and the luminance of the individual patches was varied randomly rather than being equated. This ensured (as much as possible) that neither edge artifacts nor luminance differences could be used to discriminate the target from the background; thus, the target can be detected only by color discrimination.

Stilling's pseudoisochromatic plates are all of “vanishing” design, so called because the target is read correctly by the normal trichromatic observer and not seen by the color defective (i.e., the “vanishing” figure is pseudoisochromatic to the color defective). A later variant was the “transformation” plate, where two figures are embedded in the background, one that is read by the normal observer and another that has the appropriate chromatic and lightness contrast to be read by the color-deficient observer. The Japanese ophthalmologist Ishihara published in 1917 the first edition of a set of plates that included vanishing and transformation plates, as well as hidden plates, in which the color defective can see a digit that is camouflaged for the normal by random color variation. Another design is the diagnostic or classification plate which is a more sophisticated version of the vanishing design plate that allows for differentiating between the type of red-green color deficiency, i.e., protan or deutan. The Ishihara plates are not designed for examining yellow-blue deficiencies, i.e., tritan

deficiency. In 1954, the AO-HRR (American Optical – Hardy, Rand, and Rittler) plates were printed and comprise both screening and diagnostic plates for tritan defects, as well as diagnostic and grading plates for protan and deutan defects. The plates have vanishing designs containing geometric shapes (circle, cross, and triangle) that are printed in neutral colors on a background matrix of gray dots. The saturation of the neutral colors increases in successive plates to produce designs with progressively larger color difference steps identifying different levels of deficiency.

The Ishihara and the AO-HRR pseudoisochromatic plates are often used together because their functions are complementary. The Ishihara plates are used for screening for red-green color deficiency and the AO-HRR plates to confirm protan/deutan classification, estimate the severity of red-green deficiency, and identify tritans.

Some pseudoisochromatic tests have designs containing shapes or “pathways” for the examination of nonverbal subjects and young children.

Administration of Pseudoisochromatic Plates

The majority of pseudoisochromatic tests have been designed to be viewed at 60–70 cms and standardized for natural daylight illumination (or for CIE standard illuminant C). The examiner turns the pages and controls the viewing time. An introductory plate is often included to demonstrate the visual task.

Examples of Pseudoisochromatic Plates

A number of other pseudoisochromatic tests have been produced, but none have been used as widely as the Ishihara and the AO-HRR plates. Tests have been published in Japan, e.g., the Ohkuma plates (1973), Tokyo Medical College test (1957), and Standard Pseudoisochromatic Plates (1st and 2nd editions) and in the USA, e.g., the Dvorine plates; in Germany, e.g., the Velhagen-Broschmann plates (29th edition published in 1992); in

Sweden, e.g., the Bostrom-Kugelberg plates (1972); and in France, e.g., the Lanthony Tritan Album, among many others. Despite many contenders, the pseudoisochromatic plates of Ishihara remain today the dominant instrument for routine screening of color vision.

Cross-References

- ▶ [CIE Standard Illuminants and Sources](#)
- ▶ [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- ▶ [Spectral Luminous Efficiency](#)

Psychological Color Effects

Nilgün Olguntürk

Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

Synonyms

[Emotional color effects](#); [Perceptual color effects](#)

Definition

Psychological color effects are the outcome of color sensations as determined by human perception and emotion, including color influences on feelings, thinking, and behavior.

Introduction and Classification

Colors may alter people's percepts and may affect how they feel. Colors with their psychological effects are most studied under three main areas: advancing and receding/retreating colors, in relation to temperature (warm vs. cool) and in relation to emotions. Advancing and receding/retreating

colors are related with visual perception where people tend to see some colors as if they are closer than they actually are or vice versa. The effect of color on warmth perception is related with feeling warm, which could be a physical or sensorial effect. In this case, some colors may increase or decrease the feeling of warmth, although the actual physical temperature remains the same. In relation to emotion, people's emotions may alter when they are exposed to different colors. There may be even more ways that colors may effect percepts and emotions, but the most studied three topics will be expanded in this overview.

Advancing and Receding/Retreating Colors

Colors in Isolation

In studies treating colors in isolation, Luckiesh provided one of the earliest arguments on the "retiring" and "advancing" effects of color letters placed in the same plane [1]. In 1918, he used an apparatus with red and blue filters "of fairly high purity" with which he altered the color of the letters *X* and *E* viewed inside wooden boxes. The subjects moved the red *X* until it appeared to lie in the same plane as the blue *E*. He found a lot of inter-subject variability, but still in most of the cases it was necessary to move the red *X* further away in order to make it appear to be in the same plane with the blue *E* [1, also cited in 2]. One explanation on the advancing quality of red comes from the operational mechanism of the eye. The lens of the eye has to adjust to focus the red light wavelengths, as their natural focal point lies behind the retina. Thus, red advances, creating the illusion that red objects are closer than they actually are [3]. This phenomenon is about the monocular depth perception of the eye where shortwave light refracts in the eye's optical media more than longwave light; thus, the equidistant sources of different colors cannot be simultaneously focused on the retina, which is also called chromatic aberration [4].

Successive studies on the advancing property of color found instances when blue was judged to be "nearer" than red. Pillsbury and Schaefer, in

1937, had subjects view either red neon or blue neon and argon lights [5]. When the lights were placed equidistantly, the blue light was judged the nearer. The explanation for this "conflict" in apparent nearness of red vs. blue is explained by subsequent research that suggests lightness (or brightness) to be the controlling quantity for apparent distance [2]. The Purkinje shift, a well-known physiological phenomenon, demonstrates that blue light appears lighter than red at low luminance levels.

Colors in Combination

In studies with color combinations, the relationship between perceived colors, e.g., contrast, becomes important. Mount et al. conducted the first outdoor research of color distance in 1956 [6]. They mounted four chromatic (yellow, green, red, blue) and four gray papers on plywood. The plywood with chromatic colors and grays attached on it was viewed outdoors against a dirt bank under full sun. Participants compared a chromatic color or a gray against a dark or light gray standard on each trial. It was found that each chromatic color was judged to be closer than its nearest matching gray. Each of the hues and the grays appeared closer when viewed against a dark standard rather than the light one. They found no difference in advancement for one hue over another. On the other hand, as saturation of a color was increased with respect to its background, its apparent position advanced (an effect equal at most to 1.5% of the standard distance). Colors having high lightness contrast with their background appeared advancing (an effect equal at most 3% of the standard distance). Thus, increasing relative contrast by increasing an object's lightness and/or saturation as compared to its background makes the object appear closer [2].

These findings are supported by the 1960 study of Oyama and Nanri [7]. They had participants compare standard and variable circular shapes in all combinations of achromatic and chromatic relations on varying backgrounds under laboratory conditions. They found that the apparent size of the figure increased as its lightness increased, while the lightness of the background decreased.

They found no effect of hue on apparent size [2, 7].

Egusa confirmed the findings of the studies above [8]. In 1983 he found an effect of hue when hemifields of different hues were compared for perceived depth. The green-blue difference in perceived depth was smaller than the red-green difference, with red appearing nearer. He also noted that a higher saturated color was judged nearer when it was red or green, but there was no such effect for the blue [8].

In the studies above, it appears that there may be a combined effect of all attributes of color, hue, saturation, and lightness, on advancing or receding/retreating colors, especially when they are viewed in combination. Lightness seems to be the most dominant attribute. Colors having high lightness contrast with their backgrounds appear advancing. Saturation seems to be the second most important attribute of color in judgments of nearness. As saturation of a color increases with respect to its background, it seems closer in the visual field [2].

In Relation to Temperature (Warm/Cool)

The definition of warm is directly associated with color as “having the color or tone of something that imparts heat, specifically of a hue in the range yellow through orange to red” [9]. Warmth perception is a multisensory experience of the immediate environment including physical (e.g., thermal properties, surface properties, etc.) and sensory (e.g., visual, tactile, olfactory, etc.) aspects.

Color in Relation to Physical Aspects (Perceived Temperature) of Warmth

Hue affects warmth perception independent from saturation and lightness [10, 11]. Although no relationship was found between colored illumination and perceived temperature [12], the following studies establish a relationship between applied hue (as in wall finishing colors) and perceived temperature. Itten’s study reports that in a blue-green room, participants felt cold at 15 °C, whereas they felt cold at 11.1 °C in a red-orange room [13]. In 1976 Porter and Mikellides found

that occupants preferred 2.2° higher indoor environment temperature in a blue room than in a red room. Occupants felt cold at 22.2 °C in a blue room, while this temperature was preferable in a red room. In a red room, they felt warm at 24 °C [14].

Color in Relation to Sensory Aspects of Warmth (Warm Colors Versus Cool Colors)

Warm colors are perceived warmer than cool colors [11, 15]. In interiors, one study reports that red and yellow induce the perception of warmth, while white is regarded as the coolest color [16]. In applied surface colors, hue influences warmth perception, where warm colors (red, orange, yellow) are regarded as being warmer than cool colors (green, blue, purple).

In a study on “nasal thermal sensation,” participants felt cold in their nasal nostril when smelling green-colored water and felt warm in their nasal nostril while smelling red-colored water. Thus, the visual sense affects the warmth perception of the olfactory sense as well [17].

Color Emotion

Color as an effective design tool influences people’s emotions in interior spaces. Red is one of the powerful colors and creates the highest number of emotional responses [18]. It is associated with surprise, happiness and sadness [19], energy, vitality and power [20], fear, and anger [21]. Thus, red evokes emotions in a range from the negative (anger, fear, and sadness) to positive (happiness, energy, power) ones [22].

Green is associated with nature and trees, which creates a feeling of comfort with a number of positive emotions such as feelings of relaxation, happiness, comfort, peace, and hope [23].

Blue is connoted with infinity and serenity in relation to the sky and the sea, representing peacefulness and relaxation [24]. Blue was found also to be associated with happiness [19], calmness, peacefulness, relaxation, modernism, coldness and dullness [25], and sadness [26].

Users usually desire colors in order to feel something or in order to create a psychological bond with an object or a space. It should also be

noted that if people indicate any color as being their least favorite in their daily lives, in some instances they also associate an interior with that specific color with disgust [22].

Psychosomatic Effects of Color

Each color and color combination has its own sensation. Individually or in relation to each other, colors may become eye irritants or may cause headaches. They can effect brain waves, hearth rates, blood pressure, and respiratory rates [27]. There is an ancient and widespread faith in the healing power of color, some of which is studied under color healing or color therapy [28, 29].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Mixed Chromatic Adaptation](#)
- [Unique Hues](#)

References

1. Luckiesh, M.: On "retiring" and "advancing" colors. *Am. J. Psychol.* **29**, 182–186 (1918)
2. Camgöz, N.: Effects of hue, saturation, and brightness on attention and preference. Dissertation, Bilkent University, Ankara (2000) (Also available by UMI, Bell & Howell Co., Ann Arbor, 2001)
3. Mahnke, F.H., Mahnke, R.H.: *Color and Light in Man-Made Environments*. Van Nostrand Reinhold, New York (1987)
4. Sundet, J.M.: Effects of color on perceived depth: review of experiments and evaluation of theories. *Scand. J. Psychol.* **19**, 133–143 (1978)
5. Pillsbury, W.B., Schaefer, B.R.: A note on advancing-retreating colors. *Am. J. Psychol.* **49**, 126–130 (1937)
6. Mount, G.E., Case, H.W., Sanderson, J.W., Brenner, R.: Distance judgment of colored objects. *J. Gen. Psychol.* **55**, 207–214 (1956)
7. Oyama, T., Nanri, R.: The effects of hue and brightness on the size perception. *Jpn. Psychol. Res.* **2**(1), 13–20 (1960)
8. Egusa, H.: Effects of brightness, hue and saturation on perceived depth between adjacent regions in the visual field. *Perception*. **12**(2), 167–175 (1983)
9. Merriam-Webster online dictionary. <http://www.merriam-webster.com/dictionary/warm>
10. Wright, B.: The influence of hue, lightness and saturation on apparent warmth and weight. *Am. J. Psychol.* **75**(2), 232–241 (1962)
11. Wastiels, L., Schifferstein, H.N.J., Heylighen, A., Wouters, I.: Red or rough, what makes materials warmer? *Mater. Design*. **42**, 441–449 (2012)
12. Berry, P.C.: Effects of colored illumination upon perceived temperature. *J. Appl. Psychol.* **45**, 248–250 (1961)
13. Itten, J.: *The Elements of Color*. In: Birren, F. (ed.). (trans: Van Hagen, E.). New York: Van Nostrand Reinhold (1970)
14. Porter, T., Mikellides, B.: *Colour for Architecture*. Studio Vista, London (1976)
15. Newhall, S.M.: Warmth and coolness of colors. *Psychol. Rec.* **4**, 198–212 (1941)
16. Wastiels, L., Schifferstein, H.N.J., Heylighen, A., Wouters, I.: Relating material experience to technical parameters. *Build. Environ.* **49**, 359–367 (2012)
17. Michael, G.A., Rolhion, P.: Cool colors: color-induced nasal thermal sensations. *Neurosci. Lett.* **436**, 141–144 (2008)
18. Boyatzis, C.J., Varghese, R.: Children's emotional associations with colors. *J. Gen. Psychol.* **155**(1), 77–85 (1994)
19. Terwogt, M.M., Hoeksma, J.B.: Colors and emotions. *J. Gen. Psychol.* **122**(1), 5–17 (1995)
20. Kaya, N., Crosby, M.: Color associations with different building types: an experimental study on American college students. *Color. Res. Appl.* **31**(1), 67–71 (2006)
21. da Pos, O., Green-Armytage, P.: Facial expressions, colors and basic emotions. *Color Design Creat.* **1**(1), 1–20 (2007)
22. Helvacioğlu, E.: *Color-emotion associations in interior spaces*. Dissertation, Bilkent University, Ankara (2011)
23. Kaya, N., Epps, H.N.: Relationship between color and emotion: a study of college students. *Coll. Stud. J.* **38**(3), 396–405 (2004)
24. Fehrman, K.R., Fehrman, C.: *Color: the Secret Influence*. Prentice-Hall, Upper Saddle Rive (2000)
25. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color. Res. Appl.* **32**(2), 144–150 (2007)
26. Zentner, M.R.: Preferences for colors and color-emotion combinations in early childhood. *Dev. Sci.* **4**(4), 389–398 (2001)
27. Kaiser, P.K.: Physiological response to color: a critical review. *Color. Res. Appl.* **9**(1), 29–36 (1984)
28. Helen, V.: *Colour*. Marshall Editions, London (1983)
29. O'Connor, Z.: Color psychology and color therapy: caveat emptor. *Color. Res. Appl.* **36**, 229–234f (2011)

Psychological Color Space and Color Terms

David L. Bimler

School of Psychology, Massey University,
Palmerston North, New Zealand

Synonyms

Color appearance solid; Color similarity space;
Relational color space

Definition

Psychological color space is the relational structure among color stimuli that can be found using empirical tasks that assess color similarities. Color terms are the lexical categories (which can vary across different ethnolinguistic groups) that are used to label, or describe, color appearances organized as meaningful partitions of the psychological color space.

Psychological Color Space

A color-ordering system requires three independent modes of variation to accommodate the full gamut of color experience and arrange all possible hues within a coherent continuum, in which similar pairs of colors are adjacent neighbors. This requirement, so familiar now, only emerged after a long history of trials with one- or two-dimensional schemes [1]. Here colors are considered in isolation as opposed to in the context of other colors where induced simultaneous contrast effects complicate color appearance.

“Psychological color space” is synonymous with a “color solid” containing all physically attainable hues. Numerous psychological qualities or attributes have been nominated as its underlying dimensions ([1], Chap. 5). The long delay before the acceptance of the three-dimensionality of color space suggests that most of these attributes are not immediately evident and require a

trained observer. Perhaps the most intuitive of these qualities, the circular sequence of hues around a color wheel, only became an organizing principle after Newton’s description of the prismatic spectrum. Later theorists refined the Newtonian color wheel by inserting nonspectral purples between the red and violet spectral extremes.

One tradition portrays the color solid as a sphere or as a pair of cones or pyramids joined at their bases. The “equator” is the color wheel, containing the purest, most saturated form of each hue. A *hue angle* locates each hue within the rainbow sequence. Perpendicular to the wheel and through its center runs an achromatic axis of gray tones, connecting the polar extremes of white and black. Derived forms of a given hue share the same angular coordinate, occupying space between that point on the wheel and the axis. In effect they comprise one page in a “book of color” radiating out from a central spine. In the Natural Color System – one example of a color space – these intermediate hues are derived from the pure hue by dilution with various amounts of black or white or both.

Clearly this is a cylindrical coordinate system. As well as the angular coordinate, it implies a radial coordinate (the distance out from the axis), variously labeled as “saturation” or “colorfulness” or “purity.” The third coordinate, running parallel to the axis, is some version of “lightness.”

Hering [2] argued that in addition to lightness (defined by its extremes black and white), “red” and “green” are the extremes of a second opposition. Each cancels out the other so the two cannot coexist. “Blue” and “yellow” are a third opposition. These three oppositions form an alternative way of spanning color space, with rectilinear axes which specify a hue by its content of “redness vs. greenness” and “blueness vs. yellowness.” In terms of radial coordinates, the “cardinal hues” of yellow, red, blue, and green might be located at 0°, 90°, 180°, and 270°, respectively.

In practice the purest yellow hues are relatively light, and conversely the purest purples are relatively dark. Thus, all saturated hues can only be located on the equator of the color solid by treating the lightness scale differently on each of

the radial color “pages.” The alternative is to treat “lightness” as an *objective* property that can be measured with a photometer and slice color space into planes of equal lightness, with the most saturated example of each hue located empirically in the appropriate plane (in some depictions of color space, lightness is replaced with a “brightness” axis, a more subjective quality which also increases with saturation). The effect is to tilt the color wheel to form an oblique angle with the achromatic axis, distorting the geometry of double-cone and double-pyramid models of color space.

In addition, “maximum saturation” is not a constant quality around the wheel. The most saturated “reds,” for instance, are perceptually more intense than the most saturated blues, i.e., more dissimilar from the corresponding gray tone. Incorporating these aspects of color perception produces a less regular color solid. In the widely used “Munsell color space” [3], the radial coordinate (chroma) applies the same metric across hues, reaching its highest values for the intense red hues of intermediate lightness so that they bulge out on the side of the solid. In explanations of how color names are assigned to color space, the geometrical salience of this region marks the red hues as particularly “namable,” along with “black” and “white” [4].

Munsell space emphasizes the principle of *uniformity*, that the distance between two hues in the space should reflect the perceptual dissimilarity between them. That is, a given distance at any region of color space, such as a unit increment in saturation level, should correspond to the same inter-hue dissimilarity. This empirical approach also led Munsell to expand the blue/red quadrant of the color wheel and insert “purple” as a fifth cardinal hue, contrary to the Hering polarities.

In another application of the uniformity principle, hue samples were chosen to form a square grid spanning each equal-lightness plane [5]. That is, dissimilarities were intended to be the same between each pair of adjacent samples, as far as departures from Euclidean geometry would permit. The hue samples are also specified by their coordinates on three rectilinear axes *L* (lightness), *g*, and *y*, but the *g* and *y* scales were determined

afterward rather than fundamental to this uniform color scale scheme (the OSA-UCS). Again, the resulting color solid is irregular.

An extension of the uniformity principle takes advantage of multidimensional scaling methods (MDS) to explore color ordering. Here ratings of dissimilarity among pairs of hue samples are elicited from observers and analyzed with MDS algorithms, yielding geometrical solutions where points representing the samples are arranged so that distances among them reflect the dissimilarities. In a series of empirical studies, the samples were selected from the Munsell system to vary across all three coordinates [6]. The results generally support that system’s uniformity.

Numerous other models of color space exist, emphasizing different aspects of color experience. There is a tension between the uniformity principle and other features that might be desirable, in particular the “linearity” principle that mixtures of two colored lights should be represented by points along a straight line. CIE tristimulus space is an example of a physiological model where distances are strongly nonuniform.

Given a corpus of unconstrained labeling data, in which subjects collectively applied a pattern of color terms to each stimulus, one measure of dissimilarity between any pair of stimuli is the difference between the two patterns used to describe them [7]. These dissimilarities can be reconciled with distances in a geometrical color space by introducing a “metric,” i.e., a distance/dissimilarity function that varies with location in the space [7].

Why Three Dimensions?

In the final analysis, three dimensions are necessary – though they can be conceptualized and spanned in different ways – because human color vision rests on the existence of three distinct populations of photoreceptor cells in the retina (*S*, *M*, and *L* cones), reducing the spectral information within a given stimulus to three levels of cone outputs. In the first stage of visual processing, neurons in the retina create combinations of the raw cone outputs that are more informative than single values in isolation: a sum (*L* + *M* or “luminance”) and two cone-output differences

(L-M and S-(L + M)). These combinations provide the three axes of *cone excitation space*, widely used in vision research to define experimental stimuli and results.

However, physical light can vary along far more than 3 degrees of freedom. Thus, two stimuli with physically different light spectra may appear perceptually identical (“metamers”) and be shown as the same point in color space because they provide the same stimulation to each of the cone classes. That is, trichromatic vision captures only a fraction of the information contained within the spectra of stimuli in the natural visual world. Factor analysis of spectra from a range of objects under a range of lighting conditions indicates that to capture most of the spectral variation would require six or seven distinct photopigments and additional color dimensions [8].

Notably, higher-dimensional spaces are observed in species of reptiles and birds. The three-dimensionality of human color vision suggests that additional photopigments bestowed no advantage over the course of human evolution, i.e., the differences among metamers were not significant in the ecological niches occupied by prehuman primates.

The majority of mammals make do with dichromatic color vision, having only two cone classes. This in turn implies that trichromatic vision *was* adaptive for some primates, with evolution selecting for the divergence of L- and M-photopigment genes. Several adaptive advantages have been advanced to account for this feature of human color space. The difference between M and L cones allows normal trichromats to detect red objects against green backgrounds. In particular, ripe fruit stands out against foliage, as do older, more digestible leaves [9]. Further speculative benefits from the red-green dimension of color space include social signaling, a sexual selection role, and preserving color constancy under changes in lighting conditions [8].

Abnormal Color Space

Color space is convenient for representing the effects of color vision deficiencies (CVDs).

Depending on the specific form of deficiency, two colors may be harder to discriminate (i.e., perceptually more similar) when they are separated along a particular direction in the space. This characteristic direction – running between red and green, for the common forms of CVD – is the *confusion axis*. Thresholds for distinguishing between a reference hue and small variations of it are increased in this direction (e.g., slightly redder and slightly greener versions). The locus of hues that are just noticeably different from the reference, which would form a circle when plotted in a suitable color plane, instead becomes an ellipse elongated along the confusion axis.

Alternatively, one could say that the *personal* color space for a CVD observer is compressed, reducing distances along the confusion axis. The degree of compression corresponds to the severity of the deficit. This interpretation underlies the design of widely used panel tests for CVDs. In the extreme case where an entire class of photoreceptor is absent, color space becomes two- rather than three-dimensional, and the color circle collapses to a line. The principle is supported by numerous studies where MDS reconstructed personal color spaces for individuals with CVDs.

Color Terms

As seen above, color space provides a framework for analyzing color language. For a given *Basic Color Term* (BCT), the range of hues to which that term can be ascribed (its extension) comprises the corresponding *Basic Color Category* (BCC). The BCC can be treated as a geometrical domain by mapping that range into color space.

Each BCC is circumscribed by its boundaries with adjacent color categories. “Red,” for instance, has boundaries with the domains of “pink,” “orange,” “purple,” and “black.” On one side of the “watershed,” a stimulus might be labeled more often than not as “red,” whereas a second stimulus, nearby in color space, might attract “pink” as the majority name. Thus, the continuity of color is *partitioned* into color-category domains. There are four points to note here:

1. The vicinity of a category boundary creates *categorical perception* effects in perceptual tasks. These effects are vitiated by verbal interference, indicating that language plays some role in them.
2. For English, the partition is *complete* in the sense that the category domains cover color space in its entirety; there is an applicable term for any realizable color. This is not necessarily the case in every language.
3. Empirically, the domains of English color categories are *connected* and *convex*. That is, there are no situations where two hues receive the same color term but have some hue between them in color space which is labeled with a different term. In particular, there is no term with a domain separated into two non-contiguous parts.
4. The boundaries of a BCC are part of the BCT's definition, between them circumscribing the hues that it can label. They are not *all* of its definition, however. A color term is also characterized by its *focal hues*, i.e., the hues picked out from the color gamut as prototypal best examples. One may also define the BCC "centroid" hue, the "center of mass" or mean location of its domain in color space.

So far the requirements for a color term to be "basic" have not been spelled out. One criterion have been assumed implicitly: a BCT must be understood by almost all speakers of a language, with consensus about its meaning, i.e., few disagreements or misunderstandings resulting from divergent usage. A second criterion is that to be a color term at all, a word must generalize across semantic domains and be applicable to any arbitrary object, as opposed to words like "roan," "palomino," and "dun," specific to animal coats.

In English, 11 color words are BCTs, 3 achromatic terms, and 8 chromatic terms: black, white, gray, red, pink, orange, yellow, brown, green, blue, and purple.

English also provides a rich vocabulary of *subordinate* color terms. Some of these are compounds, generated from a BCT and made more specific by modifying it with a qualifier, e.g., "dark blue" or "bluish green," or with a noun as

exemplar, e.g., "sky blue." Others are derived directly from specific objects or pigments (e.g., "lilac" or "magenta"). That is, to use them is to use an implicit metaphor ("lilac colored"). A third criterion for BCTs is that they must not be compounds or still linked to a noun and carrying its connotations. This third criterion requires some care. The BCT "orange" arose from the name of the fruit, but the metaphor has become so routine that it is now a separate word with no citric associations in the minds of speakers.

However, the distinction between basic and subordinate terms is not clear-cut. Rather, "basicness" is a continuous gradient along which can be located by examining their usage, generative properties, etc. [10].

Color terms tend to form a nested, hierarchical pattern, with a broader basic term (*hypernym*) divisible into qualified, more specific *hyponyms*. There are exceptions, color being a continuum rather than a structure of nested boxes, and a number of subordinate terms straddle the borders between BCTs. "Turquoise" and "teal" lie between "blue" and "green" and are more specific than either, though speakers may argue about their exact usage.

Primary and Secondary Terms

Among the BCTs, a distinction is sometimes drawn between *primary* and *secondary* terms. The former are the six "cardinal hues" comprising the three oppositions of the Hering opponent-hue scheme. Primary terms lend themselves to "unique hue" judgments (see chapter ► "Unique Hues"). "Unique green," for instance, is a wavelength that is seen as "green" on its own with no admixture of "blue" or "yellow." Secondary terms are derived from overlaps or intersections of primaries. Three of them – orange, purple, and pink – are the intersections between "red" and its primary neighbors "yellow," "blue," and "white," respectively. "Brown" is the overlap between "yellow" and "black," while "gray" occupies a middle ground between the achromatic extremes of "white" and "black." However, the putative primary/secondary distinction does not appear on the basicness scale as an obvious jump. In addition, it leaves unexplained why some of the terms for

primary-term intersections are basic, while others (e.g., turquoise) are subordinate.

Primary terms are featured in the *color-naming task*. Here observers characterize a series of briefly presented stimuli using a constrained list of options. A 590-nm light, for instance, might be described as “red with some yellow” (or “70% red, 30% yellow”). Repeated exposures build up the precision of the description. The procedure allows subtle nonlinearities in color vision to be quantified, such as the Bezold-Brücke effect (the changes in chromatic content of a colored light as it becomes more intense). In this context, observers are comfortable with describing hues as combinations of “red” and “yellow” when a secondary term such as “orange” is not an available response. This contrasts with the situation when a primary term such as “blue” is absent from the options; observers prefer to leave that component of the stimulus appearance unlabeled (“none of the above”).

Color vocabulary becomes more finely divided in the course of language acquisition from childhood to maturity. One might expect primary terms to be learned earlier than secondaries, but the evidence for a standard order of acquisition is equivocal [11]. Cultural factors are involved, with a historical drift toward acquiring terms at an earlier age.

Ontological Status

The ontological status of the BCTs and the categories of experience they represent are still unclear. Are the qualities of “redness” and “blueness” part of the basic genetic patrimony of every human, hardwired into human brains (as the *universalist* position has it)? To what extent is the genetic endowment modified by formative visual environment and cultural factors (specifically, the color lexicon of one’s first language)?

A number of MDS studies have elicited judgments of dissimilarities among pairs of color *terms* rather than pairs of hues. The resulting solutions, in which the BCTs are embedded in a *conceptual* color space, show no substantial differences between languages (e.g., English and Mandarin) or from perceptual MDS models [12]. Notably, color-deficient informants

internalize the normal relationships within color space such as the diametrical opposition between “red” and “green” [13]. That is, they are intellectually aware of color relationships that they have not detected from observation. MDS has also been applied to *empirical* relationships among BCTs, using the number of samples on the boundary between two BCCs as the index of their similarity.

The issue has inspired comparisons with non-English and non-European languages. It early became evident that many cultures do not emphasize hue as a topic of discussion to the same extent as English. Many languages do not distinguish “color” from other qualities of surface appearance such as lightness, sheen, or texture: depending on context, words may refer to color or some other visual effect. A notable example is the difficulty of translating putative color adjectives in the *Iliad* from Homeric Greek into consistent English equivalents. In other languages, terms may have chromatic applications but are enmeshed in a web of cultural connotations.

Systematic cross-cultural research requires a standardized palette of hue samples. Informants are presented with each sample in term and asked to identify its color. One such palette, the Munsell array, has been featured in the *World Color Survey* and the *Mesoamerican Color Survey* and in numerous studies of individual cultures (see chapter ► “[World Color Survey](#)”).

The methodology is open to critiques, by presupposing for each culture an interest in color as a concept and the presence in each language of color terms per se – neither domain specific nor carrying primarily non-chromatic meanings [14]. That is, it presupposes what it purports to discover, imposing a modern English/European conceptual framework and only capturing those aspects of color discourse that fit. Thus, results from this research tradition may underestimate the prevalence of “non-partition” languages where some colors go unlabeled.

Bearing this caveat in mind, the WCS found the number of BCTs to vary widely across languages, making finer or coarser chromatic distinctions. Thus, many languages subsume “blue” and “green” within a broad “grue” term (some “grue languages” possess a word equivalent to “purple,”

in further evidence against the primary/secondary distinction). There is, however, a cross-cultural consensus about the focal hues and “centroid hues” of the BCTs. A language may have fewer BCTs than English, but the prototypal examples of their BCCs are generally close to an English equivalent [15]. Moreover, these linguistic distinctions are *prioritized*, in the sense that some are never made unless other “earlier” distinctions have been made already (for instance, if a language marks only three color categories, these will be “dark colors,” “light colors,” and “red”). These sequence relationships define a “Berlin-Kay sequence” (see chapter ► “Berlin and Kay Theory”).

There is clearly a special status for BCTs. Several explanations for their cross-cultural nature have been advanced [4, 16]. At the same time, CP effects on color-similarity judgments vary between languages, depending on the boundaries between BCCs. These boundaries necessarily shift with the size of the color lexicon but can differ even between languages with the same number of color terms.

Cross-References

- Berlin and Kay Theory
- Unique Hues
- World Color Survey

References

1. Kuehni, R.G., Schwarz, A.: *Color Ordered: a Survey of Color Systems from Antiquity to the Present*. Oxford University Press, New York (2008)
2. Hering, E.: *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge (1957). Translated by L. M. Hurvich & D. Jameson
3. Munsell, A.H.: *Atlas of the Munsell Color System*, 2nd edn. Wadsworth-Howland, Malden (1915)
4. Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language*. **85**, 884–892 (2009)
5. MacAdam, D.L.: Uniform color scales. *J. Opt. Soc. Am.* **64**, 1691–1702 (1974)
6. Indow, T.: Multidimensional studies of Munsell color solid. *Psych. Rev.* **95**, 456–170 (1988)
7. Griffin, L.D., Mylonas, D.: Categorical colour geometry. *PLoS-One*. **14**, e0216296 (2019). <https://doi.org/10.1371/journal.pone.0216296>
8. Maloney, L.T.: Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A*. **3**, 1673–1683 (1986)
9. Sumner, P., Mollon, J.D.: Did primate trichromacy evolve for frugivory or folivory? In: Mollon, J.D., Pokorny, J., Knoblauch, K. (eds.) *Normal and Defective Colour Vision*, pp. 21–30. Oxford University Press, Oxford (2003)
10. Kerttula, S.: Relative basicness of color terms: modeling and measurement. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color*, pp. 151–169. John Benjamins, Philadelphia (2007)
11. Pitchford, N.J., Mullen, K.T.: The developmental acquisition of basic colour terms. In: Pitchford, N.J., Biggam, C.P. (eds.) *Progress in Colour Studies: Volume II. Psychological Aspects*, pp. 139–158. John Benjamins, Philadelphia (2006)
12. Moore, C.C., Romney, A.K., Hsia, T.-L.: Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English. *J. Cogn. Culture*. **2**, 1–28 (2002)
13. Shepard, R.N., Cooper, L.A.: Representation of colors in the blind, color-blind, and normally sighted. *Psych. Sci.* **3**, 97–104 (1992)
14. Saunders, B.: Revisiting basic color terms. *J. R. Anthropol. Inst.* **6**, 81–99 (2000)
15. Regier, T., Kay, P., Cook, R.S.: Universal foci and varying boundaries in linguistic color categories. In: Bara, B.G., Barsalou, L., Bucciarelli, M. (eds.) *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pp. 1827–1835. Lawrence Erlbaum Associates, Mahwah (2005)
16. Bimler, D.: Are color categories innate or internalized? Hypotheses and implications. *J. Cogn. Culture*. **5**, 265–292 (2005)

Psychological Impact of Color

- Color Psychology

Psychological Primary Colors

- Psychologically Pure Colors

Psychologically Pure Colors

Wayne Wright

Department of Philosophy, California State University, Long Beach, CA, USA

Synonyms

Psychological primary colors

Definition

Psychologically pure colors are colors that exhibit a homogeneous appearance on some dimension of perceived color. Purity of appearance is closely tied up with theoretical understandings of the three canonical color attributes of hue, saturation, and lightness. According to the dominant Hering opponent color framework, the hues red, green, blue, and yellow are distinctive and fundamental to all hue perception, in part due to the fact that they admit of pure or unique variants and all other hues (such as orange) appear as mixtures of them. Saturation and related notions are characterized essentially in terms of purity, involving judgments regarding differences in the appearance of chromatic stimuli and achromatic stimuli. Highly saturated colors appear to have low achromatic content or to be more intensely chromatic. In the achromatic domain, the lightness dimension runs from light to dark. The corresponding extreme ends of that dimension, white and black, have an appearance that wholly excludes the other, while intermediate steps between them are grays that appear to be mixtures of black and white. The concept of purity plausibly could also be usefully applied in connection with other ways of categorizing perceived color, e.g., warm versus cool and positive feeling versus negative feeling.

Overview

The concept of subjective purity figures in discussions of the three attributes of perceived color

around which the bulk of contemporary theorizing is organized: hue, lightness (value), and saturation (chroma). For hue, the apparent purity of certain hues has played a central role in the development of the dominant perspective on the structural relations among the hues: Ewald Hering's opponent colors theory [1]. In Hering's theory, hues are denominated by terms such as "red," "blue," "green," and "yellow"; note that the terms used to refer to hues from the standpoint of theorizing about color experience need not reflect the color terms or concepts employed by particular perceivers or groups of perceivers. According to Hering's conception of the hues, the four hues just named have a special status as *primary* hues. They are supposed to be psychologically basic elements out of which all hue sensations are constructed. The primary hues are hypothesized to be arranged along two independent axes, red/green and blue/yellow. This opponent structure is thought to explain phenomenological observations about the purity of some hues and the mixed appearance of others, as it allows for *unique* percepts of each primary hue and *binary* combinations of primary hues not on the same axis. Purity is the essential criterion for being a unique hue, as unique hue percepts are those that consist entirely of a single primary hue. For example, while most greens are either bluish or yellowish, unique green is the shade of green that is perceived as having no tinge of either blueness or yellowness to it. By contrast, orange is a binary hue, always perceived as some blend of redness and yellowness, while turquoise is seen as a combination of greenness and blueness. It is important to emphasize that the purity of the unique hues and the impure nature of binary hues are thought to be phenomenologically present in experience.

Regarding the unique hues, their apparent purity goes hand in hand with their perceptual salience and the privileged status of the Hering primaries as elemental hue sensations. As for binary hues, they are said to look to contain proportions of their component primaries and to not be perceived as unanalyzable, integral wholes formed from the primaries. The experimental literature on hue scaling fits with the preceding phenomenological observations and Hering's

opponent colors theory. Subjects in these studies have no little or no difficulty understanding the task instructions and can produce judgments about hue content for all hue stimuli using only the Hering primaries (in permissible combinations for nonunique hues) and which total to 100%, with unique instances of the Hering primaries excluding all other hues [2]. To be clear, there is considerable variability across subjects when it comes to which particular stimuli are designated as instances of the unique hues [3]. The binary hues seem completely unsuitable for use in a hue scaling task, and the subjective purity of, say, unique blue makes it absurd to ask subjects to judge what proportions of purple and turquoise (or red and green, for that matter) look to be in it.

The color appearance of achromatic stimuli (i.e., those with no hue content) varies along the lightness dimension between the extremes of black and white, with a gradient of shades of gray in between. Subjective purity goes along with those achromatic extrema in much the same way it does with the unique hues. The apparent purity of whiteness has long been of cultural significance, and distinctions are commonly made between pure and impure whites. The concept of absolute or pure black has also aroused much interest, and artists for centuries have consciously sought to avoid pure blacks when rendering natural scenes. That black and white have a perceptual salience akin to that of the unique hues is evident in the fact that samples of them used in the World Color Survey received by far the most selections as “best examples” of basic color terms across all participants from the 110 languages surveyed [4]. Hering took white and black to be psychologically basic (and opposed) color sensations on a par with the primary hues. He also proposed a measure of color purity stated in terms of proportions of whiteness, blackness, and primary hues perceived in a stimulus [5].

Saturation and related notions inherently are purity concepts that involve the perceived difference between a chromatic (i.e., hue-bearing) stimulus and an achromatic stimulus. Highly saturated colors appear quite vivid and more “full” of their component hue than do less saturated colors. As with hue, scaling procedures are available for

saturation, with subjects making judgments about the chromatic purity of stimuli in terms of the percentages of their total color sensations that are chromatic and achromatic [2].

While not an essential commitment of views like Hering’s that account for color phenomenology in terms of red/green, blue/yellow, and white/black axes, throughout the last 100 years, it has been common to believe or hope that those axes have a direct neurophysiological realization. Although caution is in order when allowing phenomenology to influence neurophysiological theorizing, the apparent purity and perceptual salience of the unique hues, white, and black make it natural to suppose that they must have some correspondingly distinctive neurophysiological underpinning. With respect to perceived hue, a familiar idea is that there are separate, independently operating red/green and blue/yellow neural “systems” or “channels,” the combined outputs of which account for the particular unique or binary hue appearance encountered in experience. For example, unique green is seen when the blue/yellow system is nulled and the red/green system points to green, while a purple sensation arises when the red/green and blue/yellow systems are tipped toward red and blue, respectively. It should be pointed out that this theoretical perspective need not endorse a necessary connection between any particular kind of neural activity and any particular kind of color phenomenology across all perceivers. Rather, what matters is that each individual perceiver has her own personal profile of neural responses such that the hypothesized red/green and blue/yellow systems can exhibit the suggested null and non-null activity. In any event, the idea that subjective purity has a direct neurophysiological grounding is complicated by a range of empirical findings regarding inter alia the characteristics of color-opponent processing cells in the visual system [2, 6, 7].

Despite purity being so important to the understanding of key attributes of color experience, appeals to it have not gone unquestioned. For example, in the early part of the twentieth century, Carl Stumpf rejected saturation as a basic attribute of perceived color on the grounds that it is, rather than a perceptual characteristic pertaining to the

chromatic purity of a sensation, a cognitive abstraction that involves a comparison between a perceived color and some ideal of it [5]. Saturation and its brethren clearly have had a vexed history. However, perhaps the most interesting concerns voiced about appeals to subjective color purity concern the possibility that its application has been overly restricted in treatments of hue. Such worries ultimately touch on the basic motivations for the deeply entrenched Hering opponent colors theory.

In the philosophical and empirical literatures on color, statements that echo the remarks above about the essentially mixed appearance of binary hues are widespread. However, expressions such as “pure purple” and “pure orange” can be heard in everyday conversations and are invoked in artistic and commercial settings. It is certainly easy to find paints and other products with such labels. Some scientists and philosophers have also disputed claims about the necessarily mixed appearance of all hues other than the unique Hering primaries. They contend that their own phenomenological reflection on hues such as orange and purple reveals that there are also pure variants of them, e.g., a pure purple that does not look to contain any blue or red [8, 9]. Others have objected that empirical results concerning unique hue judgments that seem to support the thesis that the binary hues have a subordinate status to the Hering primaries are an artifact of the instructions given to participants. The alternative suggestion is that a minor modification of those instructions would lead subjects to find hues such as “unique orange” and “unique turquoise” that are just as subjectively salient as the unique Hering primaries [10].

As was observed above, the form of purity relevant to the unique/binary hue distinction is perceptual (psychological), not physical or otherwise objectively assessable. This, together with the heterodoxies just noted, invites the possibility that different perceivers may have either (i) different criteria for what counts as a pure hue percept or (ii) fundamentally different kinds of structure to their color experiences that either allow or forbid the possibility of hues such as

pure orange. Both are potential sources of experimental or theoretical confusion, as well as interesting topics of study in their own right. Another idea worth considering, likely in connection with (i), is that the frequency of pronouncements regarding the mixed appearance of all purples (oranges, etc.) is due to the pervasive influence of the Hering opponent colors theory on scientific and philosophical discussions of color. The attractively tidy picture the Hering theory offers of both psychological and neurophysiological aspects of color vision certainly could affect not only how experiments are constructed and their results are interpreted but also how individuals construe their own phenomenology. It has already been pointed out that empirical difficulties confront attempts to find neurophysiological mechanisms mirroring the hypothesized opponent hue axes of Hering’s model of color phenomenology. Further research open to the possibility that purity in hue perception might not be restricted to red, green, blue, and yellow could present a serious challenge to the phenomenological tenets that make Hering’s theory such an appealing account of color perception.

At a minimum, it appears that there is a need to further sharpen the criteria for what should count as a pure color percept, if matters such as those just noted are to be resolved. Of course, what is to be made of the construct of “psychologically pure colors” depends on how psychological color is understood. That, in turn, may have important consequences for the role played by the notion of purity. For example, certain (lexical, practical, emotional, etc.) categories that are associated with color plausibly could affect or facilitate judgments of purity in various ways. Potentially in play here are distinctions that go beyond the sort of structural relations among the hues that figure in Hering’s theory, such as “dessicated” versus “moist,” “hot” versus “cold,” or “positive feeling” versus “negative feeling” colors; see Schloss and Palmer’s entry on color preference. It may therefore be appropriate to consider that subjective color purity is not univocal but rather is highly dependent on the task used to assess it and thus should be defined in each circumstance of use in

terms of the particular parameters of its empirical assessment.

Cross-References

- [Berlin and Kay Theory](#)
- [Color Preference](#)
- [Unique Hues](#)

References

1. Hering, E.: *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge, MA (1920/1964)
2. Abramov, L., Gordon, J.: Seeing unique hues. *J. Opt. Soc. Am. A*. **22**, 2143–2153 (2005)
3. Kuehni, R.: Variability in unique hue selection: a surprising phenomenon. *Color. Res. Appl.* **29**, 158–162 (2004)
4. Cook, R., Kay, P., Regier, T.: The world color survey database: history and use. In: Cohen, H., Lefebvre, C. (eds.) *Handbook of Categorisation in the Cognitive Sciences*, pp. 224–241. Elsevier, Amsterdam (2005)
5. Kuehni, R.: *Color Space and Its Divisions*. Wiley, Hoboken (2003)
6. MacLeod, D.: Into the neural maze. In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 151–178. MIT Press, Cambridge, MA (2010)
7. Malkoc, G., Kay, P., Webster, M.: Variations in normal color vision IV. Binary hues and hue scaling. *J. Opt. Soc. Am. A*. **22**, 2154–2168 (2005)
8. Koenderink, J.: *Color for the Sciences*. MIT Press, Cambridge, MA (2010)
9. Wright, W.: Reply to broadbent. *Rev. Philos. Psychol.* **2**, 629–641 (2011)
10. Jameson, K.: Where in the world color survey is the support for the Hering primaries as the basis for color categorization? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 179–202. MIT Press, Cambridge, MA (2010)

Public Lighting

- [Road Lighting](#)

Punctum Cecum

- [Blind Spot](#)

Pure Colors

- [Primary Colors](#)

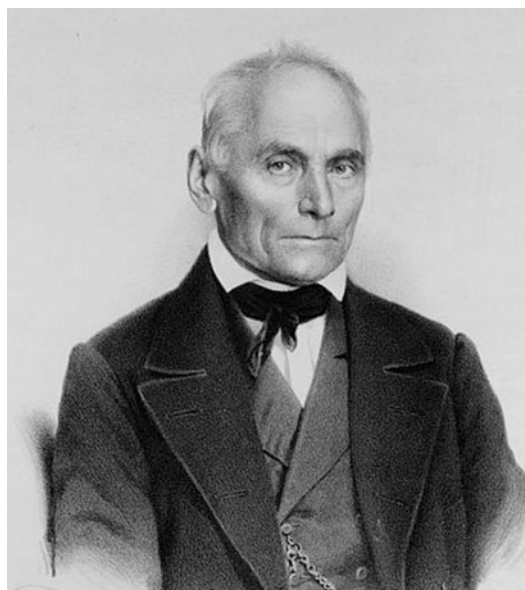
Purkyně, Jan Evangelista

Michal Vik¹ and Renzo Shamey²

¹Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic
²Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

Biography

Jan Evangelista Purkyně (Fig. 1) was born sometime between 17 December and 19 December 1787 in the castle in Libochovice near Litoměřice, Bohemia (then part of the Austrian monarchy) and now Czech Republic [1]. Purkyně died on 28 July 1869.



Purkyně, Jan Evangelista, Fig. 1 Purkyně, Jan Evangelista (1787–1869)

He was the oldest son of Josef and Rozálie (born Šafránková) Purkyně. His father worked as the chamberlain of Libochovice manor. Jan learned to observe nature and people initially from his father, who used to take him on business trips. He inherited diligence and discipline and also a sense of humor from his mother. Jan began to read early and was introduced to Komenský's publication *Orbis Pictus*. He also learned to read Latin and Greek texts from a priest called Schiffner in Libochovice.

In 1793, Purkyně's father died and Rozálie was left with her two sons Jan and Josef. With assistance from Rozálie's friends, Jan was enrolled at the Piarist Grammar school. After graduation from the grammar school (1804) and upon recommendation of his teachers, Jan went to a monastic house in Dobrá Voda. Due to his excellent performance, he managed to complete the 3-year program in 1 year. He then started to teach at the lower grammar school in Strážnice, and from there, he went to teach at the Piaristic Institute in Litomyšl. There, he devoted himself to the study of philosophy and history and poetry and became a devoted fan of the classical German philosopher F. W. Schelling.

In 1807, Jan returned home, but in the fall of the same year, he started to study philosophy in Prague University. He had to leave his studies after the third year because of financial reasons and accept to teach Baron Hildprandt's son Ferdinand in Blatná. With Hildprandt's financial support, Jan went back to Prague in 1812 to study medicine.

Upon completion of his medical program, he unsuccessfully applied to several faculty positions including professor of pharmacology in Prague, professor of anatomy and physiology in Graz, and professor of anatomy in Ljubljana.

In 1822, he secretly traveled to Germany and was introduced to Schultz and subsequently met with Goethe. In 1823, and upon the recommendation of Goethe, the Prussian king signed a decree which nominated Purkyně a professor in Physician School of Wroclaw (Breslau in Prussia) where he continued his research. There he lectured physiology, eye pathology, and psychology, but his lectures were not well attended. He

continued his study of subjective visual perception and published a comprehensive work on this topic which was dedicated to Goethe. In 1826, he was accepted as a Masonic Trainee in Berlin, 1 year later as Journeyman and subsequently as Master. In 1825, he published his work "Neue Beiträge zur Kenntnis des Sehens in Subjektiver Hinsicht" (Reimer, Berlin, pp. 109–110), which earned him the eponymy Purkinje effect (or Purkinje shift) [2].

In 1827, Jan Evangelista Purkyně (a protestant) got married to Julie Anežka Rudolphi (a catholic) in Berlin, and they kept their respective denominations. With Julie, he had two daughters and two sons. Both of his daughters died of cholera. Purkyně became a member of the Leopoldine science academy in 1829. In 1830, he became a professor of botany at Wroclaw University. In 1834, his wife died, leaving Purkyně with two young sons. He did not remarry.

Purkyně turned his focus to botany and was awarded the Montyon Prize in France for his monograph "De cellulis antherarum fibrosis nec non de granorum pollinarium formis" in 1833. In 1836, the State Secretary accepted Purkyně's suggestion for the establishment of a physiological institute (the first in Central Europe). In 1849, Purkyně was called back to Prague as a professor by the emperors' decree. Purkyně was later nominated as a member of London's King Society. Purkyně was also honored as the guardian of *Malice česká* – a committee for edification of language and literature from 1852 to 1858. Thanks to Purkyně, the first Czech industry school (1857) was also opened in Prague where he became the principal from 1857 to 1859. His house was also a place where many Czech artists met and discussed work.

Major Accomplishments/Contributions

Subjective Visual Phenomena (1818–1825)

Jan Purkyně was one of the best known anatomists and physiologists of his time. He was probably one of the early pioneers of what we now call vision science. As a medic, Purkyně studied differences in subjective visual phenomena

including lightshade, galvanic and vassal patterns (embranchment of vassals in his own eye), glare patterns, subjective feelings in darkness (phosphenes), blind spot, unity of both eyes' visual fields, double sight, indirect sight, colorblindness in peripheral retina, light patterns, and afterimages. Probably, one of the most well-known contributions of Purkyně to color and vision science is that which bears his name, the Purkyně effect.

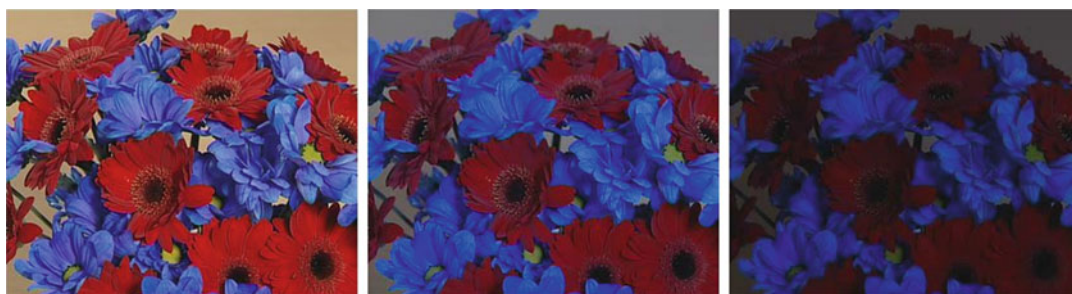
As a medical student during spring walks through the flowering countryside, Jan observed that after the sunset, the color of flowers appeared to change; red blossoms appeared darker, yellows faded, and the blues seemed brighter. Purkyně studied this phenomenon systematically and found out the same results with other object colors. That is, under decreased illumination, blue colors appeared lighter than reds. He focused his studies on this topic, and in 1818, he defended his dissertation in the faculty of medicine on "Contributions to the knowledge of sight from a subjective sense" [3]. The effect is shown in the picture of blue and red flowers simulated under three illumination conditions: average (when cones are active), dim (when rods intrude), and dark (when cones become mostly deactivated). This phenomenon is due to the contribution of rods on the perceived color of the scene. Rods require low levels of illumination to become activated (and are associated with night vision), whereas cones are activated at much higher levels of illumination (and are associated with day and color vision). Under certain conditions, e.g.,

during dawn or dusk, there is just sufficient illumination to activate cones, but illumination is low enough to activate rods also. If the contribution of rod signals to the overall image and thus the perceived color exceeds about 10%, the effect becomes known as rod intrusion. Since rods' peak sensitivity is around 496 nm, which is much closer to the peak sensitivity of the short-wavelength-sensitive cones (responsible for blue colors at 419 nm) compared to that of long-wavelength-sensitive cones (responsible for red colors at 558 nm), the overall perception of the observed scene is shifted toward blue at low levels of illumination when rods become activated. Therefore, blue colors appear brighter, whereas red colors appear very dull and almost black (Fig. 2).

Objective Eye Examination Techniques (1823)

Purkyně recommended a technique for systematic objective examination of eyes using reflective pictures. In this process, candle flame mirrors on the front and back of the cornea and then on the front and back of the retina. Purkyně examined the possibility of using reflective pictures to measure the curvature of cornea (which became the principle of keratometry and ophthalmometry) and its use in diagnosis of eye diseases and defects [4]. After many years, Purkyně obtained an achromatic Plossl microscope, which he placed in his own apartment since he did not have a suitable place within the university. He was deeply interested in the stomach's mucosa structure and discovered stomach glandule. He also observed eye

P



Purkyně, Jan Evangelista, Fig. 2 Observing *red* and *blue* flowers under average (*left*), dim (*middle*), and dark (*right*) illumination conditions. The simulation shows that

under low levels of illumination levels, the *blue* colors appear much brighter compared to *red* ones

luminescence and the possibility of observing eyes' background in vivo, a principle of ophthalmometry which was later established by Helmholtz in 1850 [5].

References

1. Bhattacharyya, K.B.: Eminent Neuroscientists: Their Lives and Works, p. 182. Bimal Kuman Dhur of Academic Publishers, Kolkata (2011)
2. J. Purkinje: Neue Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht [Observations and Experiments on the Physiology of the Senses: New Contributions to the Knowledge of Vision in Its Subjective Aspect] 2 vols., 192 pp. Georg Reimer, Berlin (1825). (in German)
3. Purkinje, J.: Beobachtungen und Versuche zur Physiologie der Sinne. Erstes Bändchen, Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht. Doctoral thesis (1819). (in German)
4. Purkinje, J.: Commentatio de examine physiologico organi visus et systematis cutanei [Contributions to physiological research of sight and skin system]. Doctoral thesis, University of Wrocław, Wrocław (1823) (in Latin)
5. Wade, N.J., Brožek, J.: Purkinje's Vision: The Dawning of Neuroscience. Lawrence Erlbaum Associates, Mahwah (2001)

Q

Quantum Dots

Ruiqian Guo, Chang Wei, Wanlu Zhang and Fengxian Xie
Engineering Research Center of Advanced Lighting Technology, Ministry of Education, Institute for Electric Light Sources, Fudan University, Shanghai, China

Synonyms

Semiconductor nanocrystals

Definition

Quantum dots are zero-dimensional nanoparticles (defined as materials of which all dimensions are smaller than 100 nm) with electrons or holes confined in three dimensions. Typically, the size of quantum dots ranges from 1 nm to 10 nm, which are made up of hundreds of atoms. Quantum dots have unique properties due to the quantum confinement effect. Their absorption and emission spectra shift to higher energy, that is, shorter wavelength, as the size of quantum dots decreases. Therefore, the photoelectronic properties can be adjusted by changing the size and composition of quantum dots.

Quantum Confinement Effect

The continuous energy band becomes quasi-continuous or discrete when the size of quantum dots is close to or less than their exciton Bohr radius. As the size decreases, the carriers, including electrons and holes, will be confined and the kinetic energy increases. Consequently, the band gap of quantum dots increases, and the absorption and emission spectra shift to higher energy, which is also called blue shift.

Classification of Quantum Dots

Quantum dots can be categorized into II–VI group (ZnS, ZnSe, CdS, CdSe, CdTe), III–V group (InP, InAs), IV–VI group (PbS, PbSe), I–III–VI group (CuInS, CuInSe), and IV group (C, Si) by their composition. Up to now, quantum dots containing Cd have been most developed such as CdS and CdSe quantum dots, which have been used in quantum dot TVs.

Mechanism of Luminescence

The luminescence of quantum dots is induced by the radiative recombination of excited electrons and holes in the semiconductor materials. Electrons in the valence band of quantum dots transit to the conduction band when excited by photons and

become “hot carriers,” leaving a hole in the valence band [1]. The hot carriers may either relax to the band edge, forming excitons (electron-hole pairs), or they may be trapped by the surface defects. Then the excitons may emit photons through radiative recombination or emit Auger electrons or phonons through nonradiative recombination.

Properties of Luminescence

- Wide excitation spectrum
- Wide tunable emission spectrum
- High color purity of emission
- High quantum yield of fluorescence
- Tunable fluorescence lifetime
- High light stability and high resistance to photobleaching (Fig. 1)

Synthesis

Quantum dots can be synthesized through physical or chemical methods. The physical methods include molecular beam epitaxy and physical vapor deposition, but the colloidal chemical synthesis is most widely used at the present time. It can be based on water or organic solvents. Furthermore, there are many factors that can affect the synthesis process, such as the temperature, reaction time, and concentration of the precursors.

There tend to be surface defects for quantum dots of small size where nonradiative recombination can happen. As a result, surface modification is used to reduce the surface defects and hence increase luminous efficiency (Fig. 2).

Energy Band Engineering

The band gap of quantum dots can be tuned through changing their size. It may work well in terms of Cd quantum dots, but it is not a panacea for all the quantum dots. Energy band engineering is thus developed these years. The band gap of quantum dots can be tuned by means of forming core-shell structure, alloying, and doping.

Applications

Quantum dots are promising in many applications such as quantum dot light-emitting diodes (QD-LEDs) [2], fluorescent probes [3], quantum dot solar cells [4], and photocatalysts [5].

Quantum Dot Light-Emitting Diodes

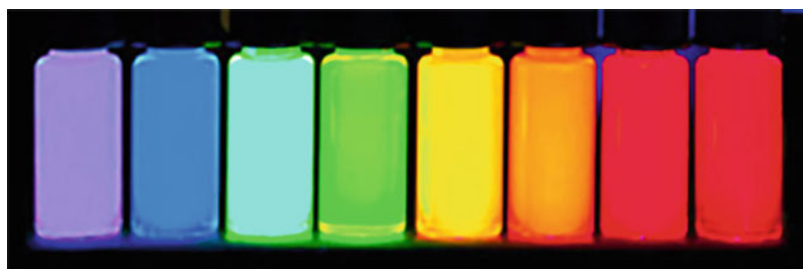
Quantum dots have been broadly used in photo-electronic devices due to their tunability, high quantum yield, and stability. Quantum dot LEDs can be divided into two types: electroluminescence and photoluminescence. In the electroluminescent devices, quantum dots are used as the emitting layer in the LEDs. Electrons and holes combine in the quantum dot emitting layer under external electric field, after passing the electron transport layer and the hole electron transport layer respectively. For photoluminescent devices, white light can be obtained by exciting quantum dots instead of traditional rare earth phosphors with UV or blue light chips (Figs. 3 and 4).

Fluorescent Probes

Quantum dots have been used as fluorescent probes in the fields of biology, medicine, and

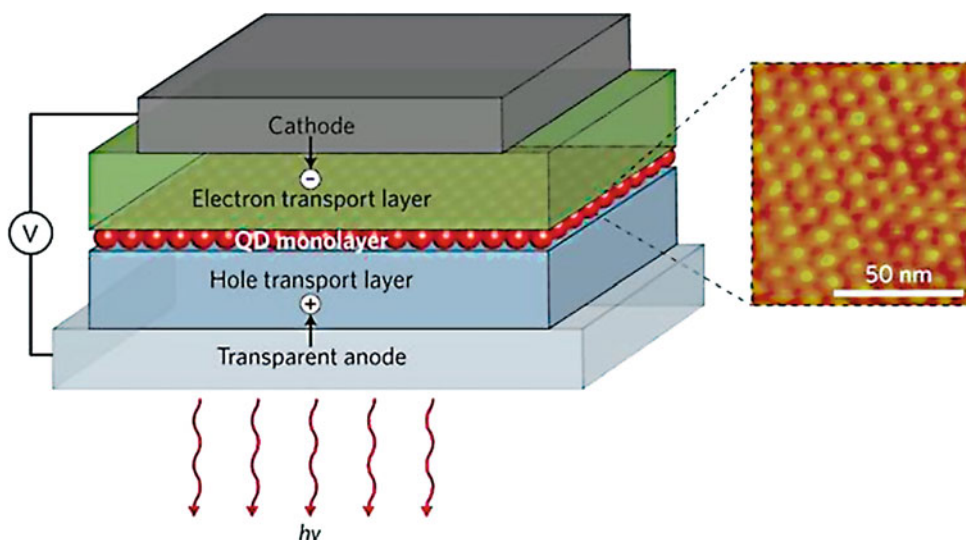
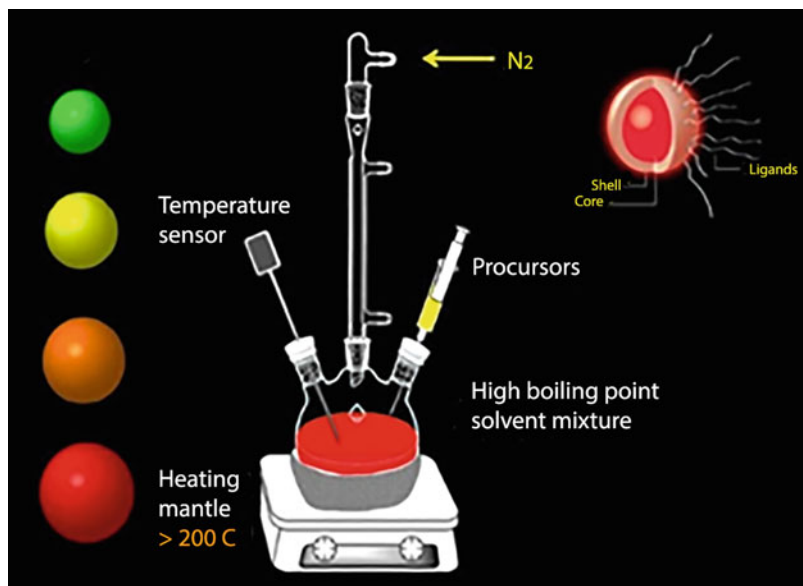
Quantum Dots,

Fig. 1 Photographs of quantum dots irradiated by UV light



Quantum Dots,

Fig. 2 Schematic representation of colloidal chemical synthesis of quantum dots



Quantum Dots, Fig. 3 Schematic representations of quantum dot LEDs [6]

analytical chemistry. For example, quantum dots of high biocompatibility can be used in in vivo imaging or real-time tracking of molecules because of their higher brightness and stability compared with traditional organic dyes. Quantum dots can also be used for detection of ions and molecules based on specific chemical reactions, fluorescence resonance energy transfer, inner filter effect, or other theories.

Quantum Dot-Sensitized Solar Cells

Typically, quantum dot-sensitized solar cells consist of photoelectrodes (e.g., titanium dioxide), conductive glass, quantum dot sensitizers, and electrolytes. Quantum dots are excited by sunlight, producing electron-hole pairs. The electrons are injected into the conduction band of the semiconductors and then to the out circuit. The oxidized electrolytes get electrons and are reduced.



Quantum Dots, Fig. 4 Photographs of quantum dot LEDs

Then the quantum dots are reduced by the reduced electrolytes and hence the original quantum dots are regenerated.

Photocatalysts

Some quantum dots, for example, ZnO, are often used as photocatalysts owing to high specific surface area and thus more active sites for photocatalysis. The quantum dots are excited by the light and produce electron-hole pairs. The electrons react with water and oxygen so hydroxyl radicals and superoxide anion radicals that can oxidize organic compounds. Rare earth ions can be doped into ZnO quantum dots to expand the active area of photocatalysis and lower the recombination rate of electron-hole pairs, which facilitates the process of photocatalysis.

Other Applications

In the fields of lasers [7], photodetectors [8], visible light communication [9], and confidential information collection [10], quantum dots have also been widely researched and utilized.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Light-Emitting Diode, OLED](#)

- [Luminescence](#)
- [Phosphors and Fluorescent Powders](#)
- [Transition Metal Ion Colors](#)
- [Visible Light Communication](#)

References

1. Pu, C., et al.: Synthetic control of exciton behavior in colloidal quantum dots. *J. Am. Chem. Soc.* **139**(9), 3302–3311 (2017)
2. Song, J., et al.: Quantum dot light-emitting diodes based on inorganic perovskite cesium lead halides (CsPbX₃). *Adv. Mater.* **27**(44), 7162–7167 (2015)
3. Hu, H., et al.: Optical sensing at the nanobiointerface of metal ion–optically-active nanocrystals. *Nanoscale*. **10**(11), 5035–5046 (2018)
4. Kamat, P.V.: Quantum dot solar cells. The next big thing in photovoltaics. *J. Phys. Chem. Lett.* **4**(6), 908–918 (2013)
5. Bajorowicz, B., et al.: Quantum dot-decorated semiconductor micro- and nanoparticles: a review of their synthesis, characterization and application in photocatalysis. *Adv. Colloid Interf. Sci.* **256**, 352–372 (2018)
6. Hirasaki, Y., et al.: Emergence of colloidal quantum-dot light-emitting technologies. *Nat. Photon.* **7**(1), 13–23 (2012)
7. Wu, J., et al.: Quantum dot optoelectronic devices: lasers, photodetectors and solar cells. *J. Phys. D: Appl. Phys.* **48**(36), 363001 (2015)
8. Ramasamy, P., et al.: All-inorganic cesium lead halide perovskite nanocrystals for photodetector applications. *Chem. Commun.* **52**(10), 2067–2070 (2016)
9. Mei, S., et al.: High-bandwidth white-light system combining a micro-LED with perovskite quantum dots for visible light communication. *ACS Appl. Mater. Interfaces*. **10**(6), 5641–5648 (2018)
10. Zhang, C., et al.: Conversion of invisible metal-organic frameworks to luminescent perovskite nanocrystals for confidential information encryption and decryption. *Nat. Commun.* **8**(1), 1138 (2017)

Quartz Halogen Lamps

- [Halogen Lamp](#)

R

Rare Earth Elements

- [Lanthanoid Ion Color](#)

Rare Earth Ions

- [Lanthanoid Ion Color](#)

Rare Earths

- [Lanthanoid Ion Color](#)

Raster Image Processing (RIP)

- [Coloration, Ink-Jet Printing](#)

Rational Models of Color Category Learning

- [Bayesian Approaches to Color Category Learning](#)

Rayleigh and Mie Scattering

David J. Lockwood
Metrology Research Centre, National Research
Council Canada, Ottawa, ON, Canada

Synonyms

[Elastic light scattering](#); [Elastic scattering](#); [Lorenz-Mie theory](#); [Lorenz-Mie-Debye theory](#); [Mie solution](#); [Mie theory](#)

Definitions

Rayleigh scattering refers primarily to the elastic scattering of light from atomic and molecular particles whose diameter is less than about one-tenth the wavelength of the incident light.

Rayleigh line refers to the unshifted central peak observed in the spectroscopic analysis of scattered light.

Mie scattering refers primarily to the elastic scattering of light from atomic and molecular particles whose diameter is larger than about the wavelength of the incident light.

Thomson scattering is elastic scattering of light from free electrons.

Raman scattering is inelastic scattering of light from objects whereby the scattered photon has a lower (Raman Stokes scattering) or higher (Raman

anti-Stokes scattering) energy than the incident photon.

Introduction

From ancient times, people have gazed up at the sky in daylight and asked the perennial question “Why is the sky blue?” [1]. Other similar and related questions are “Why is the night sky black?,” “Why are sunrises and sunsets red?,” and “Why are the clouds white?” Rayleigh [2–5] and Mie scattering [6] lie behind the long-sought answers to all such questions about the colors seen in the sky.

The phenomenon called Rayleigh scattering is named after its discoverer, John William Strutt, 3rd Baron Rayleigh (Fig. 1), a physicist who is known for a number of significant scientific discoveries that he made toward the end of the nineteenth century including the discovery of argon, an achievement for which he earned the Nobel Prize for Physics in 1904. However, the use of the term “Rayleigh scattering,” in general, and also “Rayleigh line,” in specific use in light

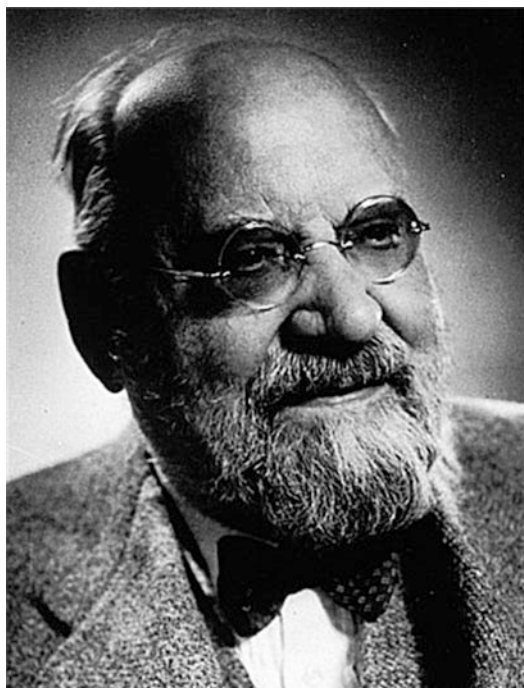
scattering spectroscopy, has been controversial in the past and has led to some confusion [7]. These nomenclature issues will be clarified in the following discussion.

It is also noteworthy that Rayleigh contributed widely to scattering theory in eight different categories; for an overview of these different contributions, see Twersky [8]. However, this entry is limited to a description of classical Rayleigh scattering of light from single small objects such as molecules and atoms. Thus, it is ignoring coherence effects that arise in solids, liquids, and gases at atmospheric pressure or even free electrons where it is known as Thomson scattering.

The scientific history behind the development of the term “Mie scattering” or “Mie theory” is convoluted [9]. Mie theory solves one of the most important problems in optics – the absorption and scattering of light by a small sphere of arbitrary size and refractive index. In 1908, Gustav Mie (Fig. 2) published his now famous paper on the simulation of the color effects connected with colloidal gold particles [[6] wherein he applied



Rayleigh and Mie Scattering, Fig. 1 Lord Rayleigh (1842–1919) ca. 1870



Rayleigh and Mie Scattering, Fig. 2 Gustav Mie (1868–1957)

Maxwell's electromagnetic theory to calculate light scattering from small spherical particles. He was not the first to perform this calculation, as similar investigations were also carried out and published earlier in 1863 by Alfred Clebsch and in 1890 by Ludvig Lorenz and soon afterwards in 1909 by Peter Debye [9]. However, Mie was the first to apply Maxwell's equations to find a solution and his name is now firmly attached to the theory, which is sometimes also called Lorenz-Mie theory and Lorenz-Mie-Debye theory.

Classical Description

Rayleigh treated the scattering of light by a spherical particle, whose diameter is much smaller than that of the wavelength of light. It is generally accepted that this size regime applies for particles (or molecules) of size less than 1/10th the vacuum wavelength, λ , of the incident light. Rayleigh considered a plane wave incident on a dielectric sphere of radius r and of relative permittivity (dielectric constant) ε , as illustrated in Fig. 3 [10]. The probability that the sphere scatters light at angle θ is proportional to the differential scattering cross section, $d\sigma(\theta)/d\Omega$, which is defined as the ratio of the power scattered into the solid angle $d\Omega$ between θ and $d\theta$ (Fig. 3) to the incident power per unit area. For unpolarized

incident light, the differential cross section becomes [10].

$$d\sigma(\theta)/d\Omega = (1/2) \times (2\pi n_0/\lambda)^4 r^6 \left[\frac{(\varepsilon - \varepsilon_0)}{(\varepsilon + 2\varepsilon_0)} \right]^2 (1 + \cos^2\theta), \quad (1)$$

where ε_0 is the relative permittivity and n_0 is the refractive index of the medium surrounding the sphere of refractive index n . Integrating this equation over the entire solid angle yields the total cross section:

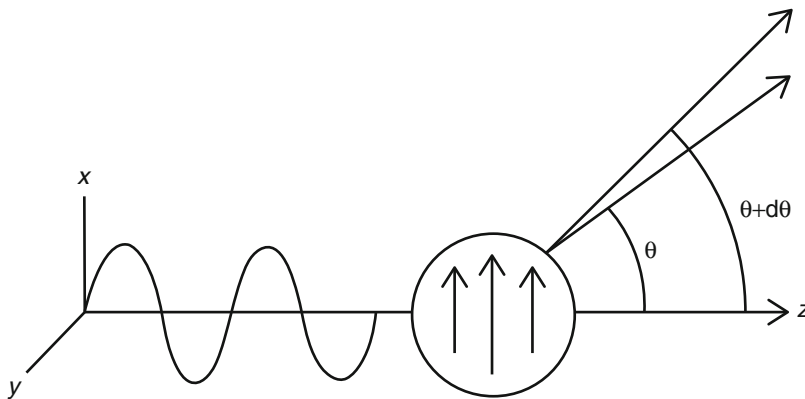
$$\sigma_R = (8\pi/3) (2\pi n_0/\lambda)^4 r^6 \left[\frac{(\varepsilon - \varepsilon_0)}{(\varepsilon + 2\varepsilon_0)} \right]^2. \quad (2)$$

The intensity I_R of the Rayleigh scattered light is given by [11]

$$I_R = I (2\pi n_0/\lambda)^4 (r^6/2D^2) \times \left[\frac{(\varepsilon - \varepsilon_0)}{(\varepsilon + 2\varepsilon_0)} \right]^2 (1 + \cos^2\theta), \quad (3)$$

where I is the intensity of the unpolarized incident light and D is the distance between the particle and the observer.

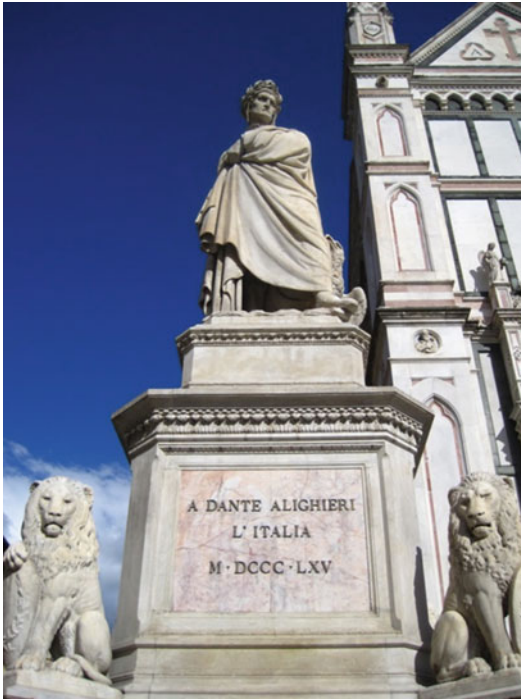
It is now possible to explain why the sky overhead looks bright blue as illustrated perfectly in Fig. 4 that shows Rayleigh scattering of sunlight



Rayleigh and Mie Scattering, Fig. 3 A plane wave polarized in the x - z plane is incident on the dielectric sphere from the left along the z direction. The arrows on the sphere denote the polarization of the dielectric material. Part of the scattered wave is shown scattered between angle

θ and $\theta + d\theta$. (Reproduced with permission from A.J. Cox, A.J. DeWeerd and J. Linden, *Am. J. Phys.* 70, 620–625 (2002). Copyright 2002, American Association of Physics Teachers)

giving rise to the intense blue color of the sky in a photograph of Dante Alighieri's effigy in Florence, Italy. Equation 4 indicates that the Rayleigh scattering intensity varies as the fourth power of the inverse wavelength of the incident light. This



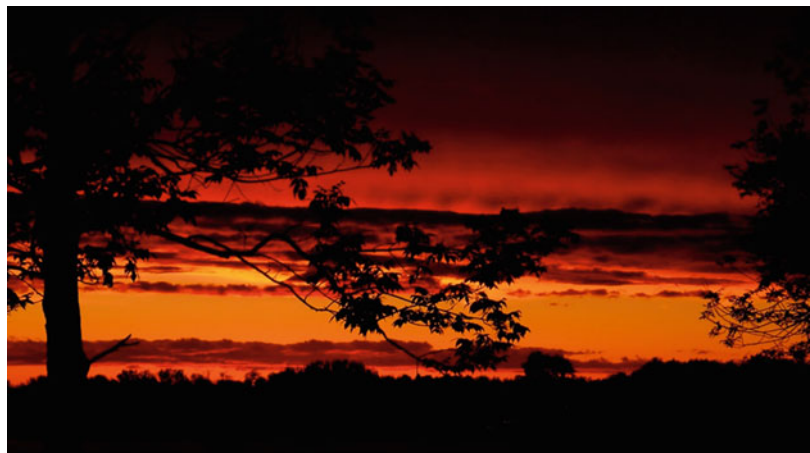
Rayleigh and Mie Scattering, Fig. 4 Dark blue skies overhead in Florence, Italy: Statue of Dante in the Piazza di Santa Croce in Florence (Enrico Pazzi, 1865). (Copyright Lilia R. Lockwood)

is known as Rayleigh's law. For example, the intensity of Rayleigh scattering at 400 nm in the blue part of the visible spectrum is higher by a factor of 9.38 compared with that at 700 nm in the red portion of the spectrum. So, when considering Rayleigh scattering of visible sunlight from molecules in the air, it is clear that blue light will be brighter than red light. Conversely, at sunset, when the sunlight travels further through the Earth's atmosphere, more blue light is scattered out of the beam path than red light, so the sunset glows red (Fig. 5). The $(1 + \cos^2 \theta)$ term in Eq. 3 contains the angular dependence of the scattering and indicates that Rayleigh scattering intensity at 90° is one-half that in the forward or backward directions.

For particle sizes larger than λ , Mie scattering predominates [6, 12], and for particles that are much larger than λ , a third type of atmospheric scattering, known as nonselective scattering, occurs [13]. A description of this last type of scattering, which can be considered as comprising a combination of Mie scattering, absorption, and multiple scattering, is outside the scope of this entry. Nonselective scattering is not wavelength dependent and is the primary cause of haze in the lower atmosphere. Water droplets and large dust particles can cause this type of scattering.

A mathematical description of the physics of Mie scattering is the result of a considerably more complicated calculation than that needed for Rayleigh scattering [12]. The calculation involves

Rayleigh and Mie Scattering, Fig. 5 This photograph of the sunset seen through clouds illustrates both Rayleigh and Mie scattering. (Copyright Terry D. Cyr)



considering a plane wave incident on the sphere and a spherical scattered wave. The total scattering cross section, σ_M , can be expressed in the form of an infinite series as [10]

$$\sigma_M = \left(\frac{\lambda^2}{2\pi n_0^2}\right) \sum (2m+1) \times \left(|a_i|^2 + |b_i|^2\right), \quad (4)$$

where the summation is from $i = 1$ to infinity and the coefficients a_i and b_i are expressed in terms of spherical Hankel functions and spherical Bessel functions of the first kind; these functions are dependent on the magnetic permeabilities of the sphere and surrounding medium and the values of parameters $m = n/n_0$ and $x = 2\pi n_0 r/\lambda$, which is termed the size parameter.

Representative results from calculations of the Rayleigh and Mie scattering cross sections using Eqs. 2 and 4, respectively, are shown in Fig. 6 in terms of the scattering efficiency, $Q = \sigma/\pi r^2$. This figure demonstrates that for small spheres with $mx \ll 1$, ($x = 2\pi n_0 r/\lambda$ and $m = n/n_0 = 1.59/1.33$), the scattering efficiencies of Rayleigh and Mie scattering are very nearly the same, whereas for larger spheres with high mx values, the behavior is completely different. The Mie scattering efficiency departs from the Rayleigh λ^{-4} behavior and approaches a limiting value of $Q = 2$, as a consequence of the “extinction paradox” related to equal

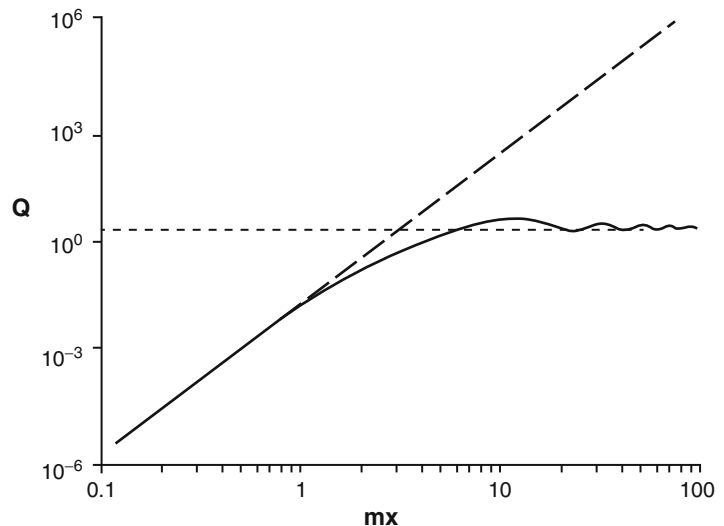
geometrical and diffraction contributions to the cross section [10]. The diffraction contribution is most readily discerned at a distance from the sphere and is strongly peaked in the forward direction. Thus, Mie scattering produces a scattering pattern like an antenna lobe with a forward lobe that becomes more intense and sharper with increasing particle size. For larger particle sizes, as Fig. 6 shows, Mie scattering is not strongly wavelength dependent. This is the reason that such scattering from water droplets in clouds, mist, or fog produces white light, as can be seen in Fig. 4 in the lower part of the sky behind the lion. Likewise, due to its larger forward lobe, Mie scattering produces the whitish glare that is seen surrounding the sun when large particulate matter is present in the air.

Quantum Description

In describing Rayleigh scattering, only the classical interaction of light with small particles has been considered so far. However, for very small particles, a proper account of Rayleigh scattering requires a quantum-mechanical description. This is given, for example, by Placzek [14] in the early days of quantum theory and, in a more modern setting, by Loudon [15], which is followed here. The scattering of light by an atom or molecule is a second-order radiative process: a quantum (or photon) of energy

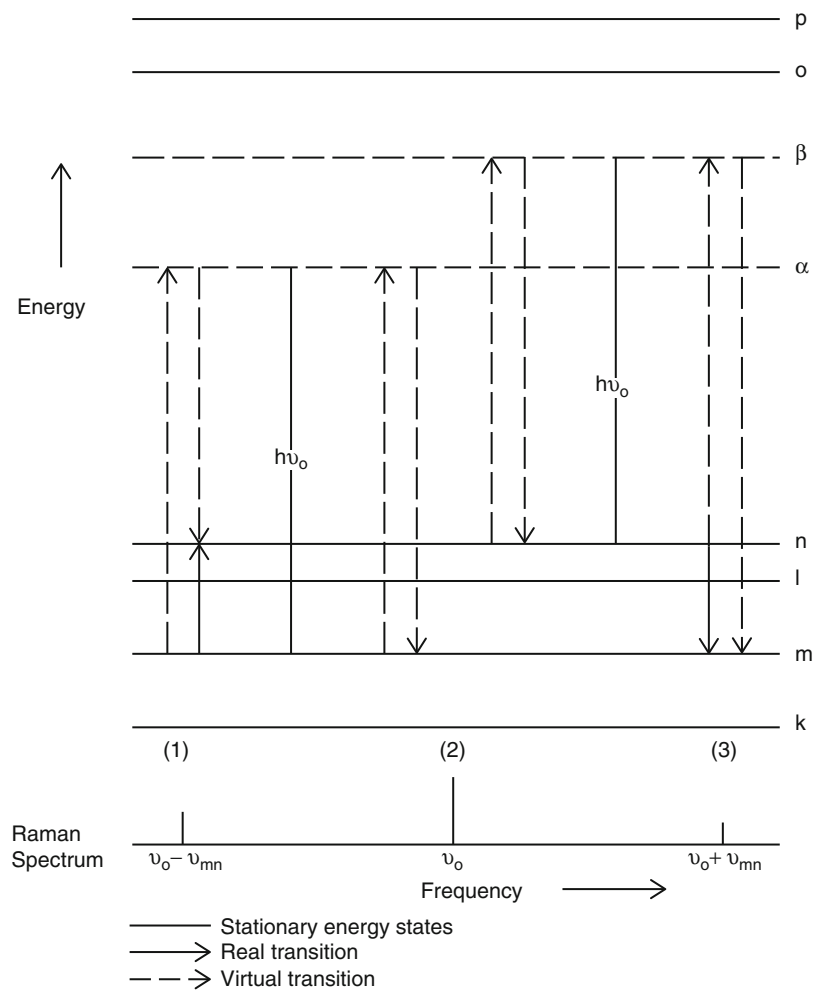
Rayleigh and Mie Scattering,

Fig. 6 Scattering efficiency $Q = \sigma/\pi r^2$ versus mx ($x = 2\pi n_0 r/\lambda$ and $m = n/n_0 = 1.59/1.33$) for Rayleigh (*dashed curve*) and Mie (*solid curve*) scattering. The dotted line indicates the limiting value of $Q = 2$. (Reproduced with permission from A.J. Cox, A.J. DeWeerd and J. Linden, Am. J. Phys. 70, 620–625 (2002). Copyright 2002, American Association of Physics Teachers)



Rayleigh and Mie Scattering,

Fig. 7 Quantum energy level transition representation of (1) Raman Stokes scattering, (2) Rayleigh scattering, and (3) Raman anti-Stokes scattering. The corresponding peaks created in the light scattering spectrum are shown below the three sets of different virtual transitions. (After David J. Lockwood, Ph.D. Thesis, Canterbury University, 1969)



$\hbar\omega$ (ω is the angular frequency and $\hbar = h/2\pi$, where h is Planck's constant) of the incident light beam is destroyed and a quantum of scattered light of energy $\hbar\omega_s$ is created. When $\omega_s = \omega$, the process is called elastic scattering or Rayleigh scattering in the nomenclature of this field of research, which is where the use of terminology leads to naming confusions for different scattering processes [7]. When $\omega_s \neq \omega$, the process is called inelastic scattering or Raman scattering; the energy difference thus created is accommodated by the atom or molecule that is involved in the scattering process. Energy may be transferred to, or taken from, the atom or molecule via an electronic transition, in which case it is called Raman Stokes or Raman anti-Stokes scattering. These three different electronic energy transition processes are illustrated in Fig. 7. Electronic

transitions can occur between any states but usually arise from or terminate at the ground state or thermally populated higher states, and the excited states are usually virtual states, that is, they are not stationary states of the system. However, if the energy of the incident or scattered light is close to or at a stationary energy state, electronic resonant scattering occurs.

After a full quantum-mechanical derivation, the differential cross section of the elastic scattering of light by an atom, $d\sigma/d\Omega$, can be written as [15]

$$d\sigma/d\Omega = \left(\frac{e^4 \omega^4}{16\pi^2 \epsilon_0^2 \hbar^2 c^4} \right) \left| \sum \{ [\mathbf{e}_s \cdot \mathbf{D}_{li} \mathbf{e}_i \cdot \mathbf{D}_{il} / (\omega_i - \omega)] + [\mathbf{e}_i \cdot \mathbf{D}_{li} \mathbf{e}_s \cdot \mathbf{D}_{il} / (\omega_i + \omega)] \} \right|^2, \quad (5)$$

for the case of the atom returning to its ground state $i = 1$ at the conclusion of the scattering event. The summation index i runs over all the eigenstates of the atomic Hamiltonian. Note that the cross section varies as the fourth power of the frequency of the incident light, as in the classical result. This equation is a simplified form of the general Kramers-Heisenberg formula for the differential cross section that also covers inelastic Raman scattering [15] and forms the basis for quantum-mechanical scattering theory. In Eq. 5, the summation is carried out over all stationary states i ; c is the velocity of light, e is the electron charge, $\mathbf{\epsilon}$ and $\mathbf{\epsilon}_s$ are the unit polarization vectors for the incident and scattered light, and $\mathbf{D}$ is the sum of the electron coordinates and is proportional to the atomic-dipole moment. Evaluating the matrix elements that determine $\mathbf{D}$ is not trivial and has only been performed for simple systems such as atomic hydrogen [16].

Equation 5 may be simplified under two limiting conditions. The first is where ω is much larger than an atomic excitation frequency ω_i , but much less than $mc^2/\hbar$ where m is the electron mass. In this high-frequency limit (corresponding to Thomson scattering), the cross section reduces simply to [15].

$$d\sigma/d\Omega = Z^2 r_e^2 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2 (\text{for } \omega \gg \omega_i) \quad (6)$$

where Z is the number of electrons in the atom and $r_e = e^2/4\pi\epsilon_0 mc^2$ is the classical electron radius. Interestingly, this result from the quantum theory for the elastic scattering cross section in the high-frequency limit is identical to that obtained from the classical theory for Thomson scattering [15].

The second limiting case is where ω is much smaller than all atomic excitation frequencies ω_i , in which case Eq. 5 becomes [15]

$$d\sigma/d\Omega = (e^4 \omega^4 / 16\pi^2 \epsilon_0^2 \hbar^2 c^4) \left| \sum (\mathbf{\epsilon}_s \cdot \mathbf{D}_{1i} \mathbf{\epsilon} \cdot \mathbf{D}_{i1} + \mathbf{\epsilon} \cdot \mathbf{D}_{1i} \mathbf{\epsilon}_s \cdot \mathbf{D}_{i1}) / \omega_i \right|^2. \quad (7)$$

This expression for the cross section in the low-frequency limit can readily be evaluated for atomic hydrogen, yielding [15, 16]

$$d\sigma/d\Omega = (81 r_e^2 / 64) \times (\omega/\omega_R)^4 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2 \quad (\text{for } \omega \ll \omega_i), \quad (8)$$

where $\hbar\omega_R = me^4/32\pi^2\epsilon_0^2\hbar^2$ is the ground-state binding energy of hydrogen.

The full expression given in Eq. 5 for the elastic scattering cross section has been determined for hydrogen by Gavrilu [16]. His tables of results for the magnitude of the scattering matrix element as a function of the frequency of the incident light have been graphically depicted by Loudon [15], and some results are illustrated in Fig. 8. For this case, the differential cross section can be written as [15].

$$d\sigma/d\Omega = r_e^2 |M_1|^2 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2, \quad (9)$$

where

$$M_1 = (m\omega^2/\hbar) \sum (2\omega_i |X_{1i}|^2) / (\omega_i^2 - \omega^2) \quad (10)$$

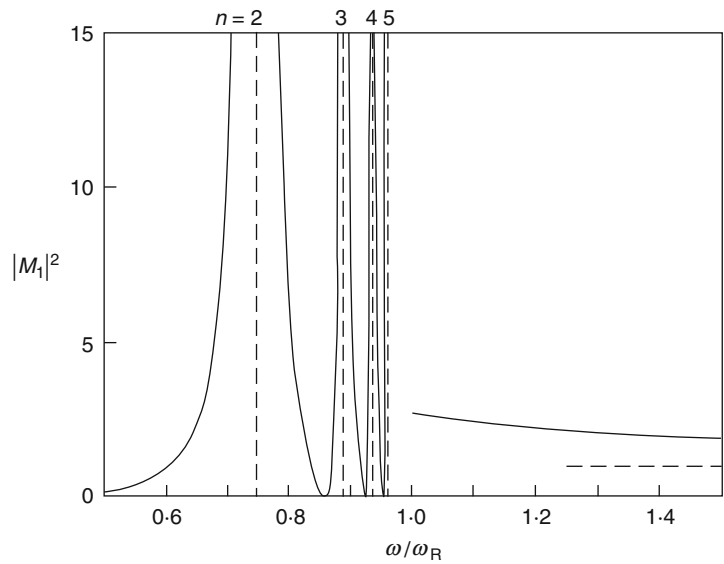
and X_{1i} are the appropriate electric-dipole matrix elements. Note that in the frequency region below ω_R , there are several strong electronic transition resonances evident (only those up to the fifth excited state were calculated by Gavrilu), separated by zeros in the differential cross section. Such resonances could not be predicted from the classical model used by Rayleigh. The inclusion of radiative damping in the theory removes the infinite values from the cross section and the accompanying zeroes evident in Fig. 8. Note also that at high frequencies, the cross-sectional value approaches that of Thomson scattering, as given by Eq. 6 with $Z = 1$, and eventually reaches it when $\omega \gg \omega_i$.

Representative Scattering Media

Here, we review Rayleigh and Mie scattering processes at work in different media of interest, including solids, liquids, and gases. Some examples of practical applications of Rayleigh and Mie

Rayleigh and Mie Scattering,

Fig. 8 Frequency dependence of the elastic scattering matrix element $|M_1|^2$ for a hydrogen atom. The vertical dashed lines show the positions of resonances with $n = 2, 3, 4$, and 5 electronic excited states. The horizontal dashed line shows the limiting value of $|M_1|^2$ for $\omega \gg \omega_R$ in the Thomson scattering region. (From Ref. [15] after the work of M. Gavrilu [16]):



Rayleigh and Mie Scattering, Fig. 9

A composite photograph of the Earth viewed from space showing white light scattering from water droplets in clouds and the surrounding blackness of space. (Copyright NASA Goddard Space Flight Center)



scattering in these representative media are also provided.

In fluids, conventional Rayleigh scattering is most commonly observed. Although proof of Rayleigh's law in gases was obtained quite early on [2, 17], at just a few isolated wavelengths, it was not until later that similar information was obtained over a wide wavelength range. In 1973,

Stone [18] reported on measurements of the Rayleigh scattering from CCl_4 and C_2Cl_4 as a function of wavelength between 600 and 1060 nm by placing the liquid sample in a hollow fused-quartz fiber and measuring the light scattered by the liquid through the fiber wall. Stone determined that the scattering loss rate was 25 dB/km for CCl_4 and 68 dB/km for C_2Cl_4 at 632.8 nm and

that the scattering loss rate followed a λ^{-4} dependence over the entire spectral range.

As mentioned in the Introduction, Rayleigh scattering is most commonly encountered in nature with the spectacular variations in color of the sky during the day and at sunset. In the absence of these scattering effects from atmospheric gases, the sky would be black, just like in outer space (Fig. 9). In practical applications, Rayleigh scattering from gaseous molecules is becoming an increasingly useful tool for diagnostic purposes. For example, Rayleigh imaging is particularly advantageous for visualizing the complex flow fields in fluid dynamics and combustion phenomena associated with shock waves and boundary layer structures having large density fluctuations [19]. In these measurements, quantitative information on these complex structures is generated by using high-power short-pulse lasers that freeze the density field in time.

In solids, there are many examples of Rayleigh scattering at work. For example, Rayleigh scattering is mainly responsible for the blue appearance of opalescent materials when viewed under white light (Fig. 10), although photonic band gap effects related to their microscopic structure frequently introduce striking color variations in their appearance, as can be seen in Fig. 10. The transmission



Rayleigh and Mie Scattering, Fig. 10 Rayleigh scattering of white light from an Australian opal. (Copyright David J. Lockwood)

of optical signals in optical fibers [20] and within microspheres or microdisks [21] is strongly affected by optical scattering processes including significant Rayleigh scattering. Glass fibers are disordered materials that exhibit microscopic variations in their density and refractive index. Rayleigh scattering from these fluctuations results in unwanted energy losses in optical transmission through the fiber with an attenuation coefficient, α_R , given by [20, 21]

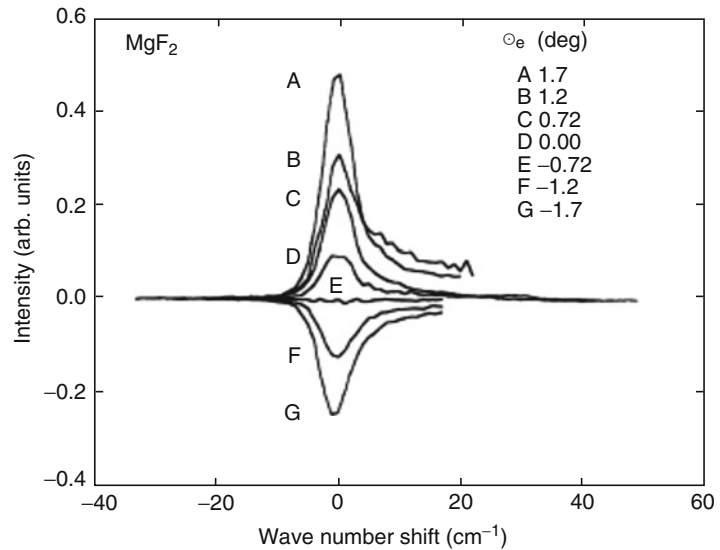
$$\alpha_R = (8\pi^3/3\lambda^4)n^8p^2kT\beta_T \quad (11)$$

where n is the fiber refractive index, p is the photoelastic coefficient of the glass, k is the Boltzmann constant, β_T is the isothermal compressibility, and T is a fictitious temperature at which the density fluctuations are “frozen” into the glass (~ 1500 K for fused silica). Note that as the contribution of Rayleigh scattering to the attenuation coefficient scales with the inverse fourth power of the wavelength, the Rayleigh scattering losses predominate at shorter wavelengths. A more unusual example is the observation of recoilless Rayleigh scattering by atoms in solids. This has been observed, for example, by employing equipment designed for Mossbauer effect studies (i.e., photon sources and analyzers with extreme selectivity in energy) to observe Rayleigh scattering in Pt, Al, graphite, and paraffin using an incident light energy of 23.8 keV [22].

Unusual Rayleigh scattering effects can also be seen in solids. For example, optical studies of gases contained in nanoporous materials with 70-nm-diameter pores show a λ^{-4} Rayleigh-type decay of the scattering coefficient with increasing wavelength [23]. Measurements of light scattering spectra from rutile structure fluorides using circularly polarized incident light have revealed Rayleigh optical activity. This occurs when the incident light is tilted away from the crystal c -axis direction of the tetragonal crystal structure [24]. Rayleigh optical activity is observed experimentally through a difference in the intensity of the Rayleigh scattering when excited with right and left circularly polarized incident light. The

Rayleigh and Mie

Scattering, Fig. 11 CID spectrum of the Rayleigh scattering in MgF_2 with the crystal rotated by angle θ_e around the laboratory y -axis (crystal a -axis). The crystal c -axis is in near alignment with the laboratory z -axis. (After Ref. [24])



Rayleigh circular intensity differential (CID), Δ_α , is defined by

$$\Delta_\alpha = (I_\alpha^R - I_\alpha^L) / (I_\alpha^R + I_\alpha^L) \quad (12)$$

where I_α^R and I_α^L are the scattered light intensities with α polarization due to right and left circularly polarized incident light, respectively. The dependence of the CID spectrum on small rotations of the crystal about the laboratory y -axis was measured in optically transparent MgF_2 and in crown glass for reference purposes. For the case of the crystal $c(a)$ axis nearly aligned along the direction of the monochromatic incident (scattered) laser light, the strong dependence of the Rayleigh CID spectrum on the crystal angle is shown in Fig. 11. Most importantly for correct identification purposes, the CID changes sign as the crystal is rotated through the true c -axis alignment at $\theta_e = -0.7^\circ$. Notably, no CID is observed for small rotations of the crystal about the orientation where the incident light and scattered light propagate perpendicular to the optic (c) axis, nor is any CID observed from the crown glass. These results indicate that the Rayleigh scattering CID observed in the MgF_2 crystal is due to depolarization effects associated with its birefringence and not polarization-dependent scattering cross sections, as confirmed theoretically [24].

Mie scattering is also found everywhere in nature: in the lower atmosphere, as noted above, in fluids like milk and latex paint, and even in biological tissue. In the latter case, Mie theory has been applied to determine if scattered light from appropriately treated tissue can be used to diagnose cancerous from healthy cells [25, 26]. Mie scattering is used in particle size determination for particles in non-absorbing media [27–29], in the determination of the oil concentration in polluted water [30], in parasitology [31], and in the design of metamaterials [32].

There has been a renewed interest recently in Mie scattering, and to a lesser extent Rayleigh scattering, for applications involving the optical properties of nanoparticles (spheres or dots) and special nanostructures such as nanorods or nanodisks. For example, the theory of Mie scattering has been extended to include the special optical properties of specific types of nanoparticles. Modeling of a Fano resonance [33] – resonances that occur in optics when a discrete localized state becomes coupled to a continuum of states – in Mie scattering from a high-index dielectric nanorod has shown that the scattering may be switched from total reflection to total transmission [34]. Remarkably, this feature means that under appropriate conditions, an incident transverse-electric-polarized light beam passes through the nanorod with practically no scattering, and the

nanorod becomes invisible for any angle of observation.

On the other hand, Mie resonances can also play a central role in presenting novel ways for boosting many optical effects near magnetic and electric multipolar resonances, as well as driving a variety of interference phenomena that govern recently discovered novel effects in meta-optics and nanophotonics [35]. Such resonances accompany the scattering of electromagnetic waves by subwavelength objects and can modify markedly the scattered light intensity and radiation pattern. Other areas of newer research related to Mie scattering include modifications of Mie theory for dielectric nanoparticles to include the toroidal dipole moment [36] as well as the usual electric-dipole moment that yields a nonradiating source (anapole) mode, which has been observed experimentally for a Si nanodisk [37], and Mie scattering color quantification has been applied in smartphone-based monitoring of microbial contamination on food products [38].

Whispering gallery mode resonators take advantage of strong light confinement and a long photon lifetime for applications in quantum optics, optomechanics, microlasers, and sensing. Cavity-enhanced Rayleigh scattering from nanoscatterers placed on the resonator surface has been shown to strongly enhance the free-space coupling of light into, and out from, the resonator [39]. Resonance Rayleigh scattering, typically involving plasmons in gold nanoparticles [40], is being applied widely to investigate biological materials such as cancer cells, food contaminants, and various metal ions (see, e.g., References [41–45]).

In summary, it is evident that Rayleigh scattering and Mie scattering are ubiquitous, being found in the everyday and colorful optical wonders that surround us.

Cross-References

- Glare
- Light, Electromagnetic Spectrum
- Maxwell, James Clerk

References

1. Lilienfeld, P.: A blue sky history. *Opt. Photon. News*. **15**(6), 32–39 (2004). <https://doi.org/10.1364/OPN.15.6.000032>
2. Strutt, J.W.: On the light from the sky, its polarization and colour. *Philos. Mag. Ser.* **4**(41), 107–120 (1871a)
3. Strutt, J.W.: On the light from the sky, its polarization and colour. *Philos. Mag. Ser.* **4**(41), 274–279 (1871b)
4. Strutt, J.W.: On the scattering of light by small particles. *Philos. Mag. Ser.* **4**(41), 447–454 (1871c)
5. Strutt, J.W.: On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *Philos. Mag. Ser.* **5**(47), 375–384 (1899)
6. Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **330**(3), 377–445 (1908)
7. Young, A.T.: Rayleigh scattering. *Phys. Today*. **35**(1), 42–48 (1982)
8. Twersky, V.: Rayleigh scattering. *Appl. Opt.* **3**, 1150–1162 (1964)
9. Hergert, W., Wriedt, T.: *The Mie Theory*. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-28738-1_2
10. Cox, A.J., DeWeerd, A.J., Linden, J.: An experiment to measure Mie and Rayleigh total scattering cross sections. *Am. J. Phys.* **70**, 620–625 (2002)
11. Hayes, W., Loudon, R.: *Scattering of Light by Crystals*, p. 2. Wiley-Interscience, New York (1978)
12. Bohren, C.F., Huffman, D.R.: *Absorption and Scattering of Light by Small Particles*. Wiley-Interscience, New York (1983)
13. Wallace, J.M., Hobbs, P.V.: *Atmospheric Science: An Introductory Survey*. Academic, Orlando (1977)
14. Placzek, G.: The Rayleigh and Raman scattering. In: Marx, E. (ed.) *Handbuch der Radiologie Part 2*, vol. 6, pp. 209–374. Akademische Verlagsgesellschaft, Leipzig (1934)
15. Loudon, R.: *The Quantum Theory of Light*. Oxford University Press, London (1973). Chapt. 11
16. Gavrilin, M.: Elastic scattering of photons by a hydrogen atom. *Phys. Rev.* **163**, 147–155 (1967)
17. Cabannes, J.: Sur la diffusion de la lumière par l'air. *C. R. Acad. Sci.* **160**, 62–63 (1915)
18. Stone, J.: Measurement of Rayleigh scattering in liquids using optical fibers. *Appl. Opt.* **12**, 1824–1827 (1973). <https://doi.org/10.1364/AO.12.001824>
19. Miles, R.B., Lempert, W.R., Forkey, J.N.: Laser Rayleigh scattering. *Meas. Sci. Technol.* **12**, R33–R51 (2001)
20. Lines, M.E.: Scattering losses in optic fiber materials. I. A new parameterization. *J. Appl. Phys.* **55**, 4052–4057 (1984). <https://doi.org/10.1063/1.332994>
21. Gorodetsky, M.L., Pryamikov, A.D., Ilchenko, V.S.: Rayleigh scattering in high-Q microspheres. *J. Opt. Soc. B.* **17**, 1051–1057 (2000)

22. Tzara, C., Barloutaud, R.: Recoilless Rayleigh scattering in solids. *Phys. Rev. Lett.* **4**, 405–406 (1960) and "Erratum" 539
23. Svensson, T., Shen, Z.: Laser spectroscopy of gas confined in nanoporous materials. *Appl. Phys. Lett.* **96**, 021107 (2010)
24. Hoffman, K.R., Yen, W.M., Lockwood, D.J., Sulewski, P.E.: Birefringence-induced vibrational Raman and Rayleigh optical activity in uniaxial crystals. *Phys. Rev. B.* **49**, 182 (1994)
25. Tsai, M.C., Tsai, T.L., Shieh, D.B., Chiu, H.T., Lee, C. Y.: Detecting HER2 on cancer cells by TiO₂ spheres Mie scattering. *Anal. Chem.* **81**(18), 7590–7596 (2009). <https://doi.org/10.1021/ac900916s>
26. Wang, M., Cao, M., Guo, Z.R., Gu, N.: Generalized multiparticle Mie modeling of light scattering by cells. *Chin. Sci. Bull.* **58**(21), 2663–2666 (2013)
27. Gompf, B., Pecha, R.: Mie scattering from a sonoluminescing bubble with high spatial and temporal resolution. *Phys. Rev. E.* **61**(5), 5253–5256 (2000)
28. Lechner, M.D.: Influence of Mie scattering on nanoparticles with different particle sizes and shapes: photometry and analytical ultracentrifugation with absorption optics. *J. Serb. Chem. Soc.* **70**(3), 361–369 (2005)
29. Brar, S.K., Verma, M.: Measurement of nanoparticles by light-scattering techniques. *Trends Anal. Chem.* **30**(1), 4–17 (2011)
30. Lindner, H., Fritz, G., Glatter, O.: Measurements on concentrated oil in water emulsions using static light scattering. *J. Colloid Interface Sci.* **242**, 239–246 (2001)
31. Serebrennikova, Y.M., Patel, J., Garcia-Rubio, L.H.: Interpretation of the ultraviolet–visible spectra of malaria parasite *Plasmodium falciparum*. *Appl. Optics.* **49**(2), 180–188 (2010)
32. Zhao, Q., Zhou, J., Zhang, F.L., Lippens, D.: Mie resonance-based dielectric metamaterials. *Mater. Today.* **12**(12), 60–69 (2009). [https://doi.org/10.1016/S1369-7021\(09\)70318-9](https://doi.org/10.1016/S1369-7021(09)70318-9)
33. Fano, U.: Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961)
34. Limonov, M.F., Rybin, M.V., Poddubny, A.N., Kivshar, Y.S.: Fano resonances in photonics. *Nat. Photonics.* **11**, 543–554 (2017). <https://doi.org/10.1038/NPHOTON.2017.142>
35. Kruk, S., Kivshar, Y.: Functional meta-optics and nanophotonics governed by Mie resonances. *ACS Photon.* **4**, 2638–2649 (2017)
36. Papasimakis, N., Fedotov, V.A., Savinov, V., Raybould, T.A., Zheludev, N.I.: Electromagnetic toroidal excitations in matter and free space. *Nat. Mater.* **15**, 263–271 (2016)
37. Miroshnichenko, A.E., Evlyukhin, A.B., Yu, Y.F., Bakker, R.M., Chipouline, A., Kuznetsov, A.I., Luk'yanchuk, B., Chichkov, B.N., Kivshar, Y.S.: Non-radiating anapole modes in dielectric nanoparticles. *Nat. Commun.* **6**, 8069 (2015)
38. Zhang, D., Liu, Q.: Biosensors and bioelectronics on smartphone for portable biochemical detection. *Biosens. Bioelectron.* **75**, 273–284 (2016)
39. Zhu, J., Özdemir, Ş.K., Yilmaz, H., Peng, B., Dong, M., Tömes, M., Carmon, T., Yang, L.: Interfacing whispering-gallery microresonators and free space light with cavity enhanced Rayleigh scattering. *Sci. Rep.* **4**, 6396 (2014)
40. Truong, P.L., Ma, X., Sim, S.J.: Resonant Rayleigh light scattering of single Au nanoparticles with different sizes and shapes. *Nanoscale.* **6**, 2307–2315 (2014)
41. Devi, R.V., Doble, M., Verma, R.S.: Nanomaterials for early detection of cancer biomarker with special emphasis on gold nanoparticles in immunoassays/sensors. *Biosens. Bioelectron.* **68**, 688–698 (2015)
42. Cai, H.-H., Pi, J., Lin, X., Li, B., Li, A., Yang, P.-H., Cai, J.: Gold nanoprobe-based resonance Rayleigh scattering assay platform: sensitive cytosensing of breast cancer cells and facile monitoring of folate receptor expression. *Biosens. Bioelectron.* **74**, 165–169 (2015)
43. Sharma, R., Ragavan, K.V., Thakur, M.S., Raghavarao, K.S.M.S.: Recent advances in nanoparticle based aptasensors for food contaminants. *Biosens. Bioelectron.* **74**, 612–627 (2015)
44. Gao, Z.F., Song, W.W., Luo, H.Q., Li, N.B.: Detection of mercury ions (II) based on non-cross-linking aggregation of double-stranded DNA modified gold nanoparticles by resonance Rayleigh scattering method. *Biosens. Bioelectron.* **65**, 360–365 (2015)
45. Chen, M., Cai, H.-H., Yang, F., Lin, D., Yang, P.-H., Cai, J.: Highly sensitive detection of chromium (III) ions by resonance Rayleigh scattering enhanced by gold nanoparticles. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **118**, 776–781 (2014)

"Red-blindness"

► [Protanopia](#)

Red-Green Deficiency

► [Deutanopia](#)

► [Protanopia](#)

Reference Standard Lamp

► [Standard Lamp](#)

Reflectance Standards

Bor-Jiunn Wen¹ and Chao-Hua Wen²

¹Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Definition

A reflectance standard is a physical reference sample including ratio values between the total amount of radiation, as of light, reflected by a surface and the total amount of radiation incident on the surface across visible spectrum. Reflectance standards are used for calibration and verification of spectrometers. There are two kinds of reflectance standards: diffuse and specular. In practice, both white and black reflectance standards are utilized for calibration. They require a similar specular or diffuse surface property to the device under test (DUT).

Applications

Diffuse and specular reflectance standards are used to calibrate colorimeters, reflectometers, spectroradiometers, bidirectional reflectance distribution function (BRDF) scatterometers [1], ultraviolet-visible (UV-VIS) spectrophotometers, and Fourier transform infrared spectroscopy (FTIR) spectrometers.

Diffuse white reflectance standard samples can be obtained with diffuse reflectance of 98% or more. Some materials can be carefully sanded (some require water with the sanding) or cleaned to refresh the surface backup to its maximum reflectance as the surface becomes soiled or contaminated. Such reflectance standards can be used for making illuminance from a luminance measurement of the standard. Moreover, with the proper measurement geometry, luminance factor

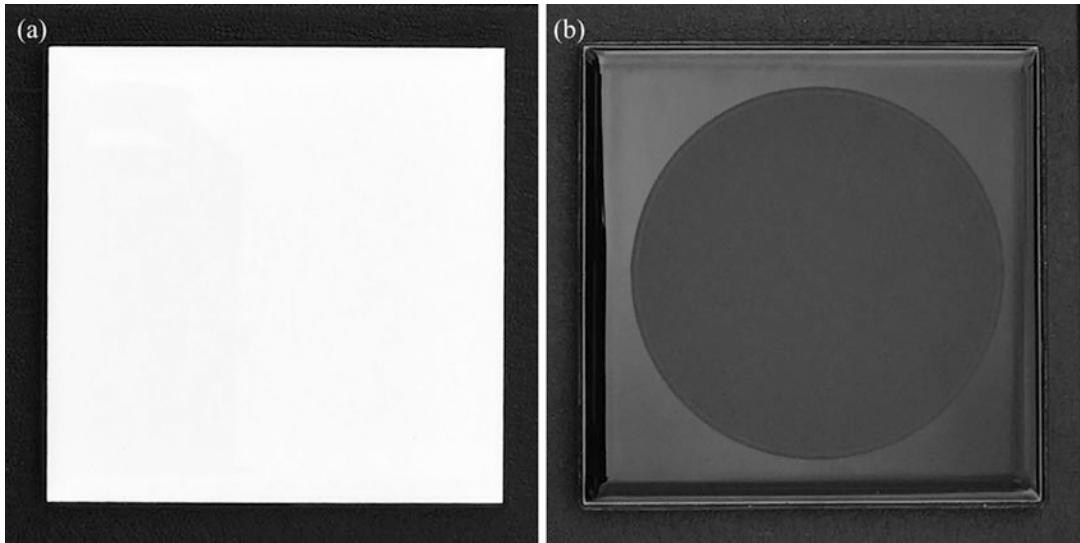
can also be obtained. Therefore, the specified measurement geometry is used to calibrate the standard. If the reflectance (or diffuse reflectance) of the standard is around 98% or 99%, it usually does refer to the reflectance, and then that value can only be used for a uniform hemispherical illumination. For industrial applications, white ceramic tiles are often used for a reflectance of 90%. If we use an isolated source at some angle, there is no reason to expect that the 99% value is even close to the proper value of the luminance factor for that geometrical configuration [1–3].

For calibration of the spectral reflection, the reference white standard and the reference black standard with the calibrated spectral reflectance factor $R_{WS}(\lambda)$ and $R_{BS}(\lambda)$, respectively, are used under the same measurement geometry [4]. Figure 1 depicts white and black specular reflectance standards, respectively. In many commercial measurement instruments, black wedges are popular as the black specular reflectance standards of optical measurements.

In order to measure the spectral reflection of the device under test (DUT), set up the measurement geometry and warm up the light source. First, measure the spectral reflectance factors of the reference white standard, $R_{WS}'(\lambda)$, with a spectrophotometer. Second, measure the spectral reflectance factors of the reference black standard, $R_{BS}'(\lambda)$. Third, measure the spectral reflectance factor of DUT, $R'(\lambda)$. The spectral reflectance factor, $R(\lambda)$, is calculated in Eq. (1):

$$R(\lambda) = [R'(\lambda) - R_{BS}'(\lambda)] \cdot \left[\frac{R_{WS}(\lambda) - R_{BS}(\lambda)}{R_{WS}'(\lambda) - R_{BS}'(\lambda)} \right] + R_{BS}(\lambda) \quad (1)$$

where $R_{WS}(\lambda)$ and $R_{BS}(\lambda)$ are the calibrated spectral reflectance factors of the reference white standard and the reference black standard, respectively. Then, the International Commission on Illumination's (CIE) tristimulus values, X , Y , and Z , are obtained as:



Reflectance Standards, Fig. 1 (a) White and (b) black specular reflectance standards

$$\begin{cases} X = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{x}(\lambda) \cdot \Delta\lambda \\ Y = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{y}(\lambda) \cdot \Delta\lambda \\ Z = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{z}(\lambda) \cdot \Delta\lambda \end{cases} \quad (2)$$

where

$$k = \frac{100}{\sum_{\lambda} S(\lambda) \cdot \bar{y}(\lambda) \cdot \Delta\lambda}$$

$S(\lambda)$: the relative spectral power distribution of the illuminant considered

$\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$: color matching functions of CIE 1931 standard colorimetric observer

$R(\lambda)$: spectral reflectance factor of the DUT

$\Delta\lambda$: wavelength interval in nm

λ : wavelength range from 380 to 780 nm

For the calibration of the filtered photometric reflection, the reference white standard and the reference black standard of calibrated CIE tristimulus values (X_{WS} , Y_{WS} , Z_{WS}) and (X_{BS} , Y_{BS} , Z_{BS}), respectively, for the same measurement geometry as in the test configuration shall be used [4]. The measuring procedures of the filtered photometric reflection of DUT are described as follows.

First, set up the measurement geometry and warm up the light source. Second, measure the CIE tristimulus values of the reference white standard, X_{WS}' , Y_{WS}' , and Z_{WS}' , with a colorimeter. Third, measure the CIE tristimulus values of the reference black standard, X_{BS}' , Y_{BS}' , and Z_{BS}' . Finally, measure the CIE tristimulus values of DUT, X' , Y' , and Z' . CIE tristimulus values of the filtered photometric reflection of DUT, X , Y , and Z , are obtained using Eq. (3):

$$\begin{cases} X = [X' - X_{BS}'] \left[\frac{X_{WS} - X_{BS}}{X_{WS}' - X_{BS}'} \right] + X_{BS} \\ Y = [Y' - Y_{BS}'] \left[\frac{Y_{WS} - Y_{BS}}{Y_{WS}' - Y_{BS}'} \right] + Y_{BS} \\ Z = [Z' - Z_{BS}'] \left[\frac{Z_{WS} - Z_{BS}}{Z_{WS}' - Z_{BS}'} \right] + Z_{BS} \end{cases} \quad (3)$$

where X_{WS} , Y_{WS} , and Z_{WS} and X_{BS} , Y_{BS} , and Z_{BS} are the calibrated CIE tristimulus values of the reference white standard and the reference black standard, respectively.

The measurement method of the spectral or the filtered photometric reflection is utilized for the reflection of DUT by using a spectrophotometer or a colorimeter. Compared to the filtered photometric reflection, results of the spectral reflection

measurement include wavelength information. However, a spectrophotometer is more expensive than a colorimeter. According to practical applications, the measurement method of the spectral or the filtered photometric reflection is appropriately utilized. Typically, color, fluorescence, grayscale, and wavelength reflectance standards are usually used in the color industry. They are used to check the reliability of operating the instruments. Various types of standards depending on different kinds of applications are introduced as following sections.

Color Reflectance Standards

Diffuse color standards are used to calibrate colorimeters and spectrophotometers in the 360–830 nm range. Compared with standards for gloss surfaces, the diffuse essential standards simplify measurements by removing the effects of viewing or illumination geometries.

Fluorescence Reflectance Standards

Fluorescence standards aid in the development of optically brightened materials such as paper and textiles and are widely used in the cosmetic industry as well. The appearance of an illuminated object is dependent upon the illumination and viewing conditions. For example, the total radiance factor of paper and board product containing fluorescent whitening agents (FWAs) is the sum of the reflected radiance factor and the luminescent radiance factor [5]. The magnitude of the luminescent radiance factor depends on the amount of UV radiation of the illumination and the fluorescence of the FWAs. It is therefore essential to calibrate not only the radiance (reflectance) factor scale but also the UV content of the illumination used in the measurement apparatus. The adjustment for the UV content is achieved by adjusting the position of the adjustable UV filter in the apparatus so that the light incident upon the sample has an effective UV content corresponding to that in the CIE illuminant C. The UV filter adjustment is based on the ISO brightness ($C/2^\circ$) value [5].

Grayscale Reflectance Standards

Grayscale diffuse reflectance standards are used to establish the linearity and accuracy of reflectance spectrophotometers and colorimeters. The standards are suitable for users who need a wide dynamic range of reflectance values that are abbreviated in range, but not in number of calibration steps. The industries typically require such a range, such as clay and processed minerals, paper, and paints, which accurately reproduces the range of reflectance values exhibited by those products.

Wavelength Reflectance Standards

Wavelength calibration standards provide stable absorption spectra for validating the wavelength calibration of spectrophotometers in the ultraviolet-visible-near-infrared (UV-VIS-NIR) region. Complete absorption spectral data are supplied with each standard. The three standards available are holmium oxide standard for the UV-VIS-NIR calibrations, dysprosium oxide standard for NIR calibrations, and erbium oxide standard for VIS-NIR calibrations. Alternatively, a mixed oxide standard is available that is doped with all three rare-earth oxides for use over the entire UV-VIS-NIR region.

Cross-References

- ▶ [CIE Tristimulus Values](#)
- ▶ [Illuminance Meter](#)
- ▶ [Instrument: Colorimeter](#)
- ▶ [Instrument: Spectrophotometer](#)
- ▶ [Standard Measurement Geometries](#)

References

1. VESA 2.0: Flat Panel Display Measurements Standard (2001)
2. CIE 015:2018: Colorimetry, 4th edn (2018)
3. SID/ICDM: Information Display Measurements Standard, v1.03 (2012)
4. SEMI D68-0719: Test Methods for Optical Properties of Electronic Paper Displays (2019)
5. ISO 2470-1:2016: Paper, board and pulps – measurement of diffuse blue reflectance factor – part 1: indoor daylight conditions (ISO brightness) (2016)

Relational Color Space

► Psychological Color Space and Color Terms

Relative Visibility Function

► Spectral Luminous Efficiency

Rendering Natural Phenomena

Oskar Elek

Computer Graphics Department, Max-Planck-Institut Informatik, Saarbrücken, Germany

Synonyms

[Simulation and image synthesis of natural phenomena](#)

Definition

In computer graphics, rendering is a process of synthetically generating an image – or a sequence of images – of an object, based on its mathematical and possibly physical description. Natural phenomena are inherently very diverse, and therefore, their rendering is a very heterogeneous field with different paradigms and approaches.

As the name implies the objects and phenomena of interest will mainly have a natural origin. However, due to similar characteristics of the underlying problems, rendering of artificial objects made of natural materials is often considered a part of the field as well.

Although the main target of rendering is the creation of images, it is usually not trivial to obtain the mathematical and physical description of the simulated entities (i.e., the input data). Because of this the rendering algorithms can potentially require coupling with simulation

methods, which provide means to computationally generate the required data. Alternatively, acquired or even hand-modeled data can be used, in case the simulation proves to be infeasible for some reason (e.g., too time-consuming).

Categories of Natural Phenomena

The richness of the natural environment of course implies the existence of a tremendous amount of phenomena and objects. The structure of this system is hardly apparent, but can roughly be classified according to the physical characteristics of the phenomena as follows (also see examples in Fig. 1):

- **Sparse phenomena**, involving gases, aerosols, and vapors but also effects of electromagnetism or high-energy particles. These include various well-known meteorological phenomena like atmospheric scattering, clouds, fog, rainbows, or lightning; astronomical phenomena such as auroras, stars, nebulae, and other stellar bodies; and also smaller-scale phenomena like fire, smoke, and dust.
- **Fluid phenomena**, most notably oceans and other large water bodies and effects associated with their surfaces, such as waves and streams. Also medium- and small-scale liquid substances, especially beverages like milk, fruit juices, and coffee; suspensions like blood, paints, and inks; and volcanic phenomena such as lava flow. In addition, especially from the perspective of simulation methodology, it is possible to regard fine-grained solid materials like sand and partly solid substances such as gels, as fluid phenomena.
- **Solid objects and phenomena**. On a large scale, primarily geological formations ranging from mountains to entire planets. On a medium scale, organic entities like vegetation, biological tissues, hair, and fur and inorganic objects such as ice and rock formations, crystals, metals and their alloys, and man-made objects manufactured from these. Among small-scale solid objects, most effort has been focused on rendering precious gems. Additionally,



Rendering Natural Phenomena, Fig. 1 Examples of sparse, fluid, and solid phenomena

rendering of natural and artificial solid objects exhibiting a layered structure has attracted research attention, for example, coated or painted objects, oxidized and patinated metals, composite materials, gemstones, and many others.

From the perspective of rendering and also modeling, another important distinction can be made between phenomena and objects with well-definable geometry and opaque surfaces and those with a dominating volumetric character (either from the spatial or the optical point of view).

The first group will likely be modeled using geometric primitives and rendered with algorithms suitable for distinguishing between discrete parts of the phenomenon, e.g., the object-air interface. The light interaction will primarily be taking place on these interfaces, mainly

as reflection and refraction. Most opaque solids belong into this category but also some fluid substances, such as clear liquids.

On the other hand, the second group will likely utilize a volumetric representation (or a combination with a geometric one) and naturally also volume-rendering methods. Materials with these properties are called participating media, and the dominating optical interactions here will be scattering and absorption. These characteristics are inherent to virtually all sparse phenomena but also to most fluid and many solid ones.

This distinction is naturally relative and without a sharp transition. As an example, one can imagine that a small stone grain will be translucent and have an apparent volumetric structure, while a large piece will appear opaque and will interact with light mostly by diffuse surface reflection (Fig. 2).



Rendering Natural Phenomena, Fig. 2 A small quartz crystal and a large piece of sandstone

Causes of Color

Another important aspect of natural phenomena from the perspective of color science is how they interact with light and produce color and the overall appearance [1]:

- **Emission**, due to *incandescence* (e.g., stellar radiation, volcanic activity, lightning), *gas excitation* (auroras, artificial gas lamps), or a combination of these (fire). The variation of color is primarily caused by varying energy density. Virtually all natural light originates in these processes.
- **Geometric causes**, such as *scattering* (atmosphere, clouds, smoke, milk, and generally almost all substances exhibiting volumetric properties or diffuse reflection), *dispersion* (rainbows, sundogs, snow), *diffraction* (opals, thin filament-like objects such as spider webs and hair, carapaces of certain bugs), *interference* (single- or multilayered structures such as bubbles and certain gemstones and insects), and *polarization* (specular reflection from smooth surfaces). These interactions are usually elastic, i.e., they conserve energy, and the color variation is mainly caused by geometric configurations and relations.
- **Absorption and reemission**, mostly in *organic compounds* (plant and animal tissues, natural and artificial pigments, dyes, and inks)

and all *metals* but also in certain minerals and semiconductors and occasionally in light liquids, most notably in *pure water*. These interactions are usually inelastic, leading to energy loss described by the Beer-Bouguer law. Color variation is caused by relative efficiency of the reemission in different wavelengths.

Historical Overview

Synthesizing physically plausible images of natural phenomena has arguably been one of the primary foci of rendering from the beginning [2]. However, the complexity of most natural phenomena prevented their plausible rendering until the early 1980s, mostly because of the lack of theoretical understanding and computational power.

A significant milestone was reached by introducing radiosity [3] and stochastic ray tracing [4] algorithms in 1984, enabling physically based simulation of global illumination effects. A more general solution was then presented in 1986 [5] by James Kajiya in the form of the rendering equation – a unified mathematical framework that enables the simulation of all effects that conform to geometric optics. These approaches were then extended to rendering of volumetric phenomena in the late 1980s and early 1990s.

Algorithmic improvements have gone hand in hand with the evolution of computing hardware,

making rendering of image sequences and even entire movies feasible by the end of the 1980s. However, probably the biggest breakthrough came with the introduction of parallel programmable graphics processor units (GPUs) to the consumer market in 2001. Their programmability enables researchers and developers to produce specialized algorithms focused on simulating and rendering effects of diverse nature. Programmable GPUs quickly became widespread and enabled the development of algorithms and applications capable of simulating numerous natural phenomena, including video games and other interactive applications.

Rendering Methodologies

The traditional view is to split (realistic) rendering as a field into noninteractive (“offline”) rendering [2] and interactive rendering [6], with the former putting emphasis on quality and the latter on speed. This applies to rendering of natural phenomena just as well. The spectrum of possible approaches is however much more continuous than that and rather depends on the *paradigm* chosen for designing a particular rendering method (refer to Figs. 3 and 4).

Empirical or phenomenological methods are generally *top-down*; their design starts by collecting observations of a phenomenon to be simulated and then proceeds to invent an arbitrary algorithm which best reproduces the phenomenon while keeping computational costs as low as possible. As such, empirical methods are used mainly in interactive applications where the important factors are visual plausibility, robustness, and speed, while physical plausibility and radiometric accuracy are secondary. Notable examples include:

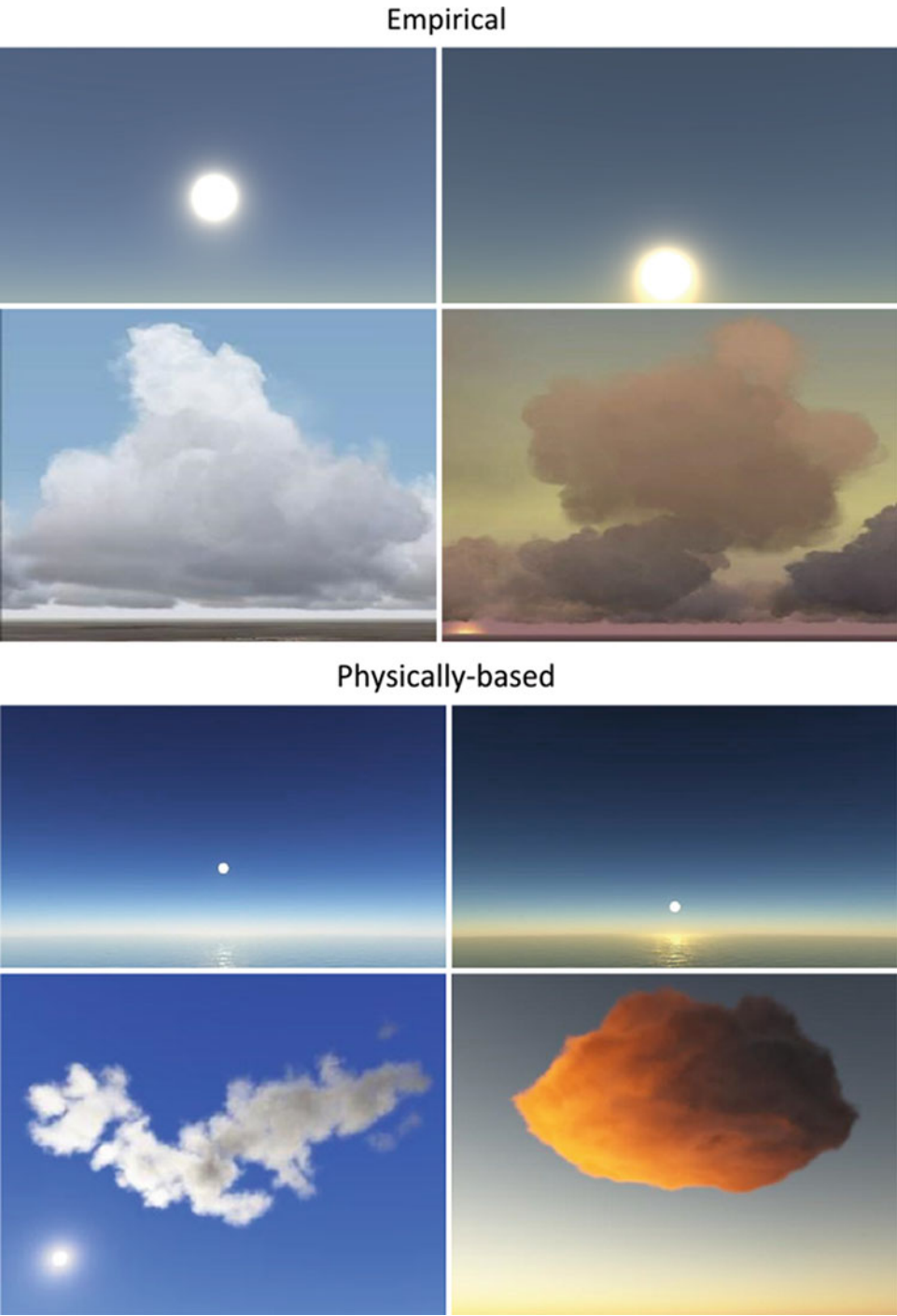
- **The Lambert reflectance model** (1760) states that reflection from a sufficiently rough surface is perfectly diffuse, i.e., isotropic with regard to observation direction. Despite building on physically meaningful assumptions (multiple scattering underneath the surface) and the fact that many materials (such as uncoated paper)

behave very closely to the model, there is no ideal diffuse reflector.

- **The Phong reflectance model** (1975) approximates reflection from glossy surfaces by the cosine function raised to an exponent proportional to the surface smoothness. This leads to reflection in the form of a blurry circular patch which gets brighter and smaller with increasing smoothness and in the limit case corresponds to the Dirac function for a perfect mirror. Despite having no physical basis, the model has been widely adopted and is still used, mainly thanks to its simplicity.
- **Particle systems** (1983) are ubiquitously used in both interactive and offline applications to model volumetric phenomena (such as clouds). In this sense, particles are small semitransparent entities often modeled by mapping a texture onto a rectangular geometric primitive that always faces the observer. If used in sufficient numbers, they are able to mask their discrete nature, while still being cheaper than the corresponding full volumetric representation (e.g., a 3D voxel grid). However, simulating global illumination effects in conjunction with particle systems requires additional effort, which might hinder their utilization. As a result it is often necessary to design another empirical model to compute illumination in an intended way, increasing the necessary amount of work.

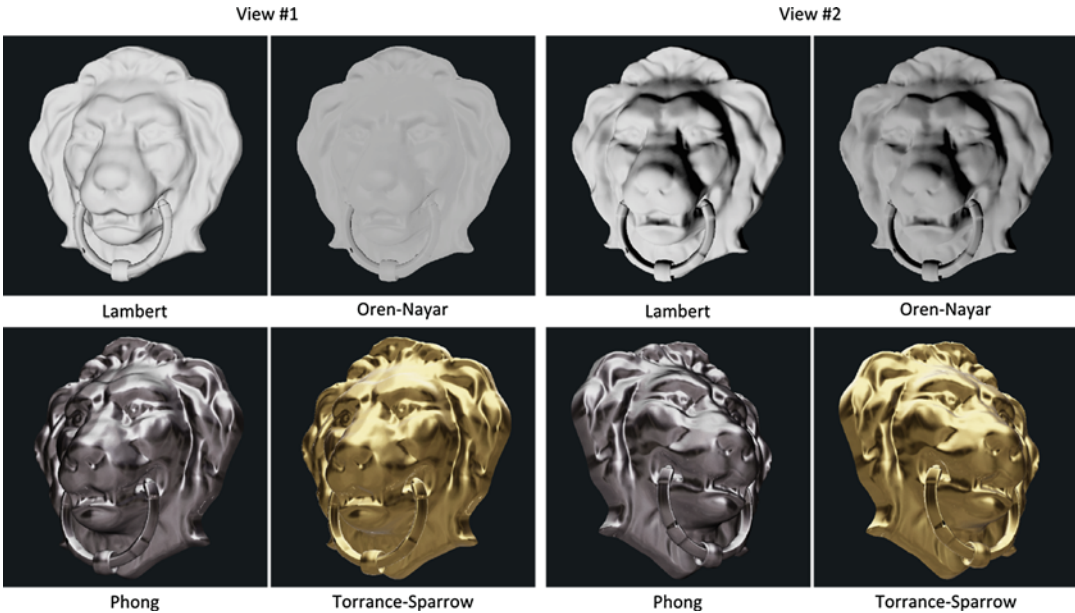
The main downsides of empirical methods are that they often require a lot of artistic supervision during the content creation process and that they seldom work outside the range of phenomena they were designed for.

Physically based methods proceed in the opposite, bottom-up fashion. Using physically meaningful data, these methods apply the laws of physics to render the target phenomenon. Consequently, rather than being centered around a particular simulated phenomenon, these methods share more principal similarities and are usually useful for rendering a whole class of phenomena to which the implemented laws and assumptions apply. On the downside, their computational cost is often much higher than that in empirical approaches. This is usually treated by applying



Rendering Natural Phenomena, Fig. 3 A comparison of interactive (empirical versus physically based) methods for rendering atmospheric scattering and clouds. The empirical approaches are typically faster, but produce less

dynamic and believable results and in some cases require additional artistic input (Image credit: Petr Kellnhofer using the model of Sean O’Neil, (*top ‘sunset’ pair*), Niniane Wang (*top ‘clouds’ pair*))



Rendering Natural Phenomena, Fig. 4 A comparison of empirical and physically based reflectance models. *Top row:* The Lambert and the (physically based) Oren-Nayar models on a diffuse gray clay material. The Oren-Nayar model correctly produces a “flatter” distribution, which is caused by backscattering on the rough clay surface. *Bottom row:* A glossy object made of gold, reflecting an environment map. The Torrance-Sparrow model

produces a more plausible reflection, which is mainly visible in grazing angles where the Phong model incorrectly produces darker reflections in disagreement with the Fresnel law. In addition, the Phong model fails to reproduce the yellow hue of gold, since it does not handle conductors properly. This could only be obtained by additional manual tweaking

Occam’s razor, simplifying the rendering algorithm in ways that decrease the simulation time but do not negatively affect the result. Examples of such methods include:

- **The Torrance-Sparrow reflectance model** (1992) simulates reflection from rough glossy surfaces (and as such being an alternative to the Phong model). The method considers the simulated surface to consist of microscopic facets oriented according to a certain statistical distribution (i.e., Gaussian distribution). Each facet is assumed to reflect light according to the Fresnel law, and additionally, probabilistic shadowing of neighboring facets is taken into account. Although in some situations the model produces results similar to the Phong model, it is energy conserving, correctly handles objects made of conductors, and behaves plausibly in grazing angles, albeit being somewhat more difficult to understand.

- **Photon mapping** (1996) [7] is a general framework for rendering global illumination effects. It is applicable to both surface and volume illumination, which makes it especially suitable for rendering natural phenomena. Its main idea lies in distributing the light energy by shooting and tracing small energy particles, photons. Every interaction of a photon with the simulated environment is recorded, and when all illumination energy is distributed, the algorithm calculates the energy density at observed locations by performing local photon density estimation. Much work has been invested into improving the original technique, resulting in one of the most versatile global illumination algorithms to date.

Originally, physically based methods were only used in offline applications. Recently, however, with the increasing performance and flexibility of the programmable hardware, the

boundary between offline and interactive methods has gotten weaker. Many physically based rendering algorithms (such as ray tracing or photon mapping) can today be implemented to run interactively, albeit with decreased rendering quality.

Finally, **predictive methods** represent a step further than physically based methods. Opposite to empirical approaches, their primary focus is physical correctness and radiometric accuracy. The utilized algorithms must be spectral, unbiased, and support all significant physical phenomena that occur in the simulated environment. They require measured and acquired input data, and the results usually need to be viewed under controlled conditions. The resulting methods typically need validation and are generally very slow, even compared to physically based approaches. As such they are useful mainly in virtual prototyping applications such as in automotive industry, architecture, and gem processing.

Phenomena Simulation

As mentioned in the “Introduction,” in addition to rendering a phenomenon (and thereby producing an image of its momentary appearance), it is often necessary to obtain data about its spatial and temporal development prior to the rendering. Virtually all natural phenomena are dynamic in a certain sense, and this dynamicity can be regarded in the short and the long term.

Short-term development usually stems from the character of the phenomenon itself and its internal dynamics. Its nature can be *synthetic* (e.g., condensation of water vapor leading to cloud formation), *evolutionary* (for instance, flow of liquid particles in a stream), or *destructive* (such as cracking or shattering of an iceberg). Capturing this behavior will often entail using discrete particle or continuous dynamics.

On the other hand, **long-term development** is usually related to interactions of a phenomenon with its surrounding environment [8], sometimes referred to as *weathering* (in case the process is destructive). Many forms of this behavior exist, for instance, terrain formation, soil erosion and

cracking, oxidation and patination of metals, organic tissue decomposition, and others.

From the methodological point of view, approaches for simulating natural phenomena can again be divided into empirical and physically based. Very often, however, a combination of these two shows to be the best compromise. The cause of this is arguably the fact that while it is often very difficult to devise an empirical method that captures the entire complexity of a given phenomenon well, a complete physically based simulation of any larger system can easily become intractable. Consequently, many approaches simulate the global behavior of the target phenomenon in a physical, possibly simplified way (for instance, using a reduced resolution of the simulated structure) and then add the remaining details empirically. The most frequent way to perform the latter is using *fractal* functions or systems:

- **Terrain rendering** methods usually generate the overall terrain morphology by simulating the orogenetic and erosive processes (or simply use satellite data) and add more detailed features by random fractal perturbations.
- **Rendering of plants** frequently employs the so-called L-systems to generate the plants and trees. L-systems are iterated functions described by grammars that imitate branching in real plants. Adding random perturbations into the system can produce plausibly looking plants with unlimited number of variants that retain the overall character defined by the L-system.
- **Rendering of ocean waves** can generate larger-scale waves with a fluid dynamics simulation or synthesizing them from trochoidal wave theory using measured wave frequency spectra and then again add random fractal perturbations to break disturbing repetitive patterns.

Despite the convenience of fractal functions, their inherent *iterative* nature makes them difficult to use in some cases. Therefore, efforts have been made to design regular functions with fractal properties. The most successful method to do so is the *Perlin noise* [9], which has become a



Rendering Natural Phenomena, Fig. 5 Multiple superposed frequencies (octaves) of Perlin noise (*top left*) and examples of procedurally generated content – terrain,

forest, and fire (Image credit: Franck Doassans (forest), Sergei Bolisov (fire))

cornerstone for generating many diverse natural phenomena and also fuelling future development in this area (see examples in Fig. 5). Since Perlin noise behaves as a regular function, it enables on-the-fly evaluation without requiring additional storage, making it feasible for use even in interactive applications.

It might of course be possible to acquire the data needed for rendering a given phenomenon, although in some cases this can prove difficult or impractical for quantitative reasons:

- The phenomenon of interest might be too large and as a result produce overwhelming amounts of data. A good example is acquisition of terrains, which, although today possible via satellite scanning, still does not produce data with resolution sufficient for some applications.
- Dynamic phenomena are generally difficult to acquire, especially in cases when the scanning process is slower than the rate of the phenomenon's significant change. Acquisition of flames and smoke is a good example here.
- Scanning processes yield limited amounts of instances of the target phenomenon. If many

such instances are needed (e.g., in cloud rendering), the cost of the acquisition process might become prohibitive.

Similar to physically, based simulation, however, these problems can in some cases be overcome by applying procedural perturbation techniques to the scanned data.

Future

The present knowledge in physics theoretically allows us to explain and hence simulate virtually all observable natural phenomena. The limiting factors in doing so are therefore always computational resources. In the future, the increasing memory density will enable us to work with larger natural systems, and the growing parallel computing power of CPUs and GPUs will allow simulation and rendering of more complex phenomena. Especially in interactive applications, the current trend of using physically based approaches over empirical ones will most likely continue.

Cross-References

- [Blackbody and Blackbody Radiation](#)
- [Global Illumination](#)
- [High Dynamic Range Imaging](#)
- [Iridescence \(Goniochromism\)](#)
- [Kubelka, Paul](#)
- [Rayleigh and Mie Scattering](#)
- [Spectral Power Distribution](#)

References

1. Nassau, K.: The Physics and Chemistry of Color, 2nd edn. Wiley-Interscience, ISBN 0471391069 Hoboken, New Jersey, USA (2001)
2. Pharr, M., Humphreys, G.: Physically Based Rendering, 2nd edn. Morgan Kaufmann, ISBN 0123750792 Burlington, Massachusetts, USA (2010)
3. Goral, C.M., Torrance, K.E., Greenberg, D.P., Battaile, B.: Modeling the interaction of light between diffuse surfaces. SIGGRAPH Comput. Graph. **18**, 213–222 (1984)
4. Cook, R.L., Porter, T., Carpenter, L.: Distributed ray tracing. SIGGRAPH. Comput. Graph. **18**, 137–145 (1984)
5. Kajiya, J.T.: The rendering equation. SIGGRAPH. Comput. Graph. **20**, 143–150 (1986)
6. Akenine-Möller, T., Haines, E., Hoffman, N.: Real-Time Rendering, 3rd edn. AK Peters, ISBN 1568814240 Natick, Massachusetts, USA (2008)
7. Jensen, H. W.: Global illumination using photon maps. In: Proceedings of EGWR, pp. 91–100 Porto, Portugal (1996)
8. Dorsey, J., Rushmeier, H., Sillion, F.: Digital Modeling of Material Appearance. Morgan Kaufmann, ISBN 0122211812 Burlington, Massachusetts, USA (2007)
9. Perlin, K.: An image synthesizer. SIGGRAPH. Comput. Graph. **19**, 287–296 (1985)

Resist Printing

- [Batik](#)

Retinal Art

- [Optical Art](#)

Retinex Theory

John McCann

McCann Imaging, Arlington, MA, USA

Synonyms

[Color and lightness constancy](#); [Land Retinex theory](#)

Definition

Retinex is the theory of human color vision proposed by Edwin Land to account for color sensations in real scenes. Color constancy experiments showed that color does not correlate with receptor responses. In real scenes, the content of the entire image controls appearances. A triplet of L, M, S cone responses can appear any color. Land coined the word “Retinex” (the contraction of retina and cortex) to identify the spatial image processing responsible for color constancy. Further, he showed that color sensations are predicted by three apparent lightnesses observed in long-, middle-, and short-wave illumination. Retinex is also used as the name of computer algorithms that mimic vision’s spatial interactions to calculate the apparent lightnesses observed in complex scenes.

Overview

Edwin H. Land, the inventor of hundreds of photographic film patents, was struck by experiments showing that color sensations in real complex images depend on scene content. Film responds to the light falling on each tiny local region. Land realized that vision’s mechanisms were very different from film. His early experiments studied the colors observed in red and white projections [1]. He realized color appearance required both the cone responses to a local region and the neural spatial processing of the rest of the scene. He proposed the Retinex Theory.

Land coined the word Retinex to describe three independent spatial channels. In 1964, he wrote: “We would propose that all of the receptors with maximum sensitivity to the long-waves in the spectrum, for example, operate as a unit to form a complete record of long-wave stimuli from objects being observed. For convenience of reference, let us call this suggested retinal-cerebral system a ‘retinex’” [2–5]. It is the word that describes the mechanisms that perform the comparison of scene information to create the array of sensations of L, M, S lightness in each of human vision’s three channels.

Cone Quanta Catch

Visible light falls on objects that reflect some of it to the eye. Color vision depends on the spectrum of the illumination falling on an object and the spectrum of its reflectance. The product of these spectra describes the light coming to the eye. There are three types of cones in normal observers that are called L for long-wave-, M for middle-wave-, and S for short-wave-sensitive cones. The receptors’ spectral sensitivities multiplied by the light falling on the retina determine the L, M, S cone responses, namely, the “quanta catch” of the cones. The cones convert the quanta catch to

nerve signals that pass through many spatial comparisons in the visual system. The cone quanta catch is the important transition from the physics of light to the physiology of vision. However, it is just the first step in the process.

Figure 1 illustrates a laboratory experiment that generates equal L, M, S quanta catches from different reflectance papers.

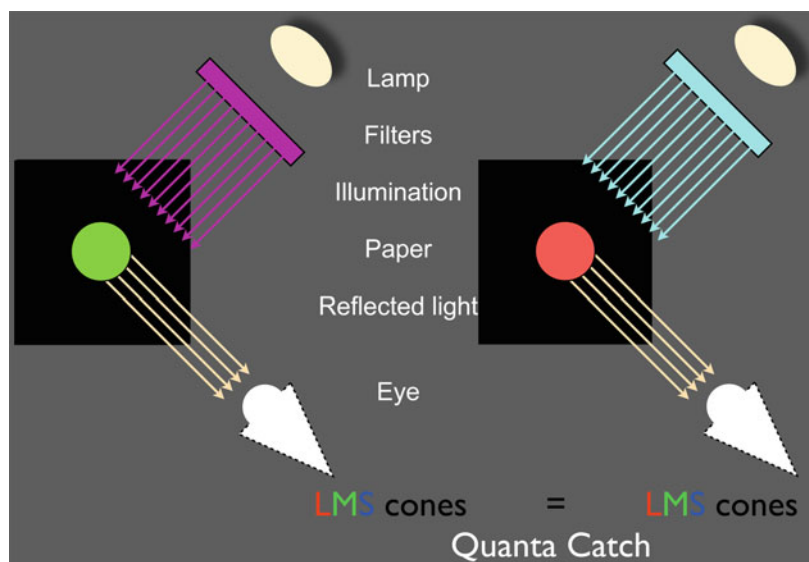
At the top of Fig. 1, light passes through filters that determine the spectrum of each illumination falling on circular pieces of paper. The light coming to the eye from the papers is modified again by the reflectance spectra of the papers. Further, in this experiment, there is no light coming from the black surrounds.

On the left, there is a tungsten light source at the top. There is a filter that absorbs more middle-wave than long- and short-wave visible light (magenta arrows). The green circular paper reflects more middle-wave than long- and short-wave light. The light coming to the eye is the product of these spectra. The integrals of that light using the three cone spectral sensitivities determine the L, M, S quanta catch values.

On the right, there is the same tungsten light source at the top. There is a different filter that absorbs more long-wave than middle- and short-wave light (cyan arrows). The right-red paper reflects more long-wave than middle- and short-

Retinex Theory,

Fig. 1 Equal quanta catches from red and green papers. In a darkroom with a black surround, identical quanta catches appear the same – neither red nor green color appearances. When viewed alone, in a no-light background, the appearance of a spot of light correlates with cone quanta catch



wave light. In this experiment, the spectra of the two illuminants and the two reflectances were adjusted to generate the same triplet of LMS cone quanta catches. Under these conditions, the left-green paper and the right-red papers are identical retinal stimuli. They appear equal to each other, but that color is neither red nor green.

Mondrian Experiments

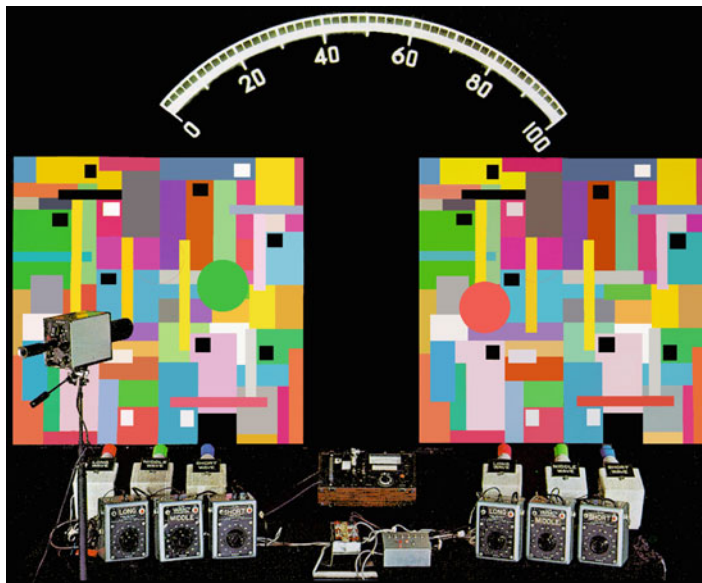
Land's color Mondrian experiment is similar, with the exception that he used a complex array of papers to simulate real-world scenes. The important difference is that in complex scenes, a particular quanta catch can appear any color: red, green, blue, yellow, white, or black.

Figure 2 shows the double color Mondrian experiment [6]. It used two identical Mondrians made of color papers, and two identical sets of illuminators. Each set had three different, non-overlapping spectral illuminants (long-, middle-, and short-wave visible light). Each projector

provided spatially uniform illumination for one spectral band, for one of the Mondrians. The amount of illumination from each projector was independently adjusted.

In this experiment, observers reported the colors of papers in the Mondrians. In this illustration, we will look at the circular red and green papers. Land adjusted the illuminant mixtures of light from the two sets of three projectors in order to control the light coming from the green and red papers. The same amounts of L, M, S light came from the green circle on the left as from the red circle on the right. First, he turned on just the long-wave lights. He adjusted the amounts of spatially uniform illumination falling on each Mondrian separately. He set the illumination so that the left-green circle and right-red circle had equal light-meter readings. Then, he did the same for middle- and short-wave light. The left-green and right-red circles had equal quanta catches by the L, M, S cones.

In this complex scene, observers reported that equal quanta catches circles appeared green on the



Retinex Theory, Fig. 2 Land's double Mondrian experiment. Two identical sets of matte colored papers, with separate L, M, S illuminating projectors with voltage transformers for control of the amount of spatially uniform light falling on each Mondrian. Telephotometer readings were projected above the Mondrians. Land separately measured

the L, M, S radiances from a green circle in the left Mondrian. Then, he adjusted the L, M, S radiances from a red circle on the right Mondrian to be the same. Observers reported different red and green colors produced by identical light stimuli

left and red on the right. Observers reported color constancy, namely, the red paper looked red and the green paper looked green despite the identical cone quanta catches.

Land repeated this experiment with all the Mondrian papers. A constant L, M, S quanta catch could generate any color sensation. The presence of the complex scene introduced more information to the visual system. The red and green papers appeared equal in Fig. 1: neither red, nor green. The red paper looked red and the green paper looked green in Fig. 2. The scene's spatial content stimulated vision's spatial image processing mechanisms to generate color constancy. The postreceptor visual processing plays a dominant role in color appearance in real scenes. Land's word "Retinex" gave this spatial process a name. He proposed a theoretical mechanism as well.

Retinex Mechanism

Figure 3 illustrates the pair of Mondrians in only long-wave light with more illumination on the left Mondrian, than on the right. The appearances of the two Mondrians are nearly constant. This is a common observation: Humans are insensitive to large changes in uniform illumination.

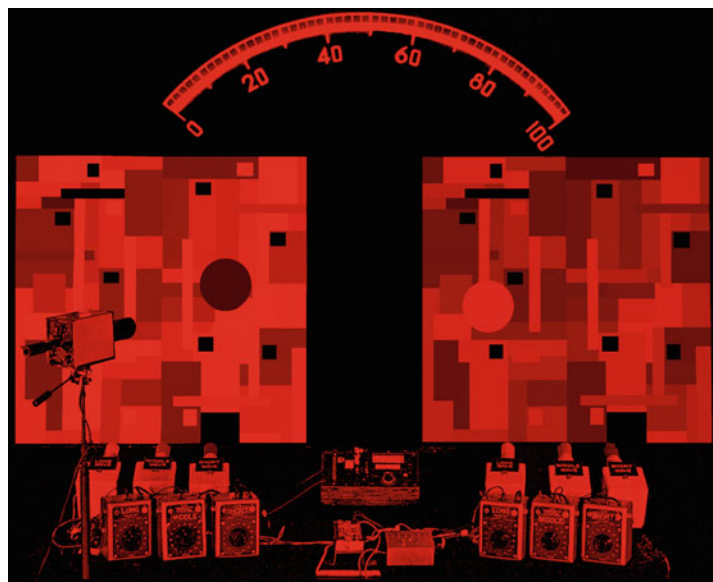
Retinex Theory,

Fig. 3 Pair of Mondrians in only long-wave light. Each rectangle in the left Mondrian has the identical reflectance as the corresponding rectangle in the right Mondrian. Only the circles have different reflectances. The left Mondrian has more illumination than the right. Observers report that the left Mondrian rectangles appear slightly lighter than the right rectangles. Each corresponding rectangle is nearly the same lightness in the left and right Mondrians

As illustrated here, in long-wave illumination, the left-green paper circle appeared dark when it generated the same L cone quanta catch as the lighter right-red circle. Vision's spatial image processing rendered the red and green papers with different apparent lightnesses in long-wave light. The lightnesses are stable with large changes in overall illumination.

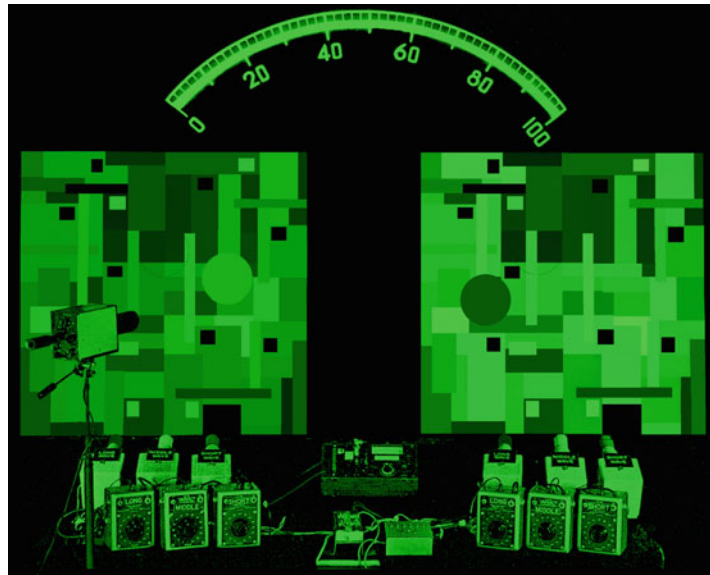
Figure 4 illustrates the Mondrian viewed in middle-wave illumination. The left-green circle now appeared light, and the right-red circle looked dark. In Fig. 4, the circles had the same M cone quanta catch. Vision's spatial image processing rendered the red and green papers with different, and opposite, lightnesses in Fig. 3 and Fig. 4.

The left-green circular paper has more long-wave and less middle-wave illumination. The right-red circular paper has less long-wave and more middle-wave illumination. When adjusted using the telephotometer, those illuminations make the red and green papers to have identical radiances. Using the same procedure, Land adjusted the short-wave illuminants to make the left and right circles to have equal radiances. Having set the L, M, S illuminations so that both circles have equal radiances, Land turned on all six projectors. Although both circles caused identical cone quanta catches, observers reported that they had very different color appearances: red and



Retinex Theory,

Fig. 4 Pair of Mondrians in only middle-wave light. Now, the left Mondrian has less illumination than the right. In this waveband, the pattern of lightnesses differs from that in long-wave light. Observers report that the right Mondrian rectangles are slightly lighter than the left ones. Each corresponding rectangle is nearly the same lightness in the left and right Mondrians



green (illustrated in Fig. 2). Using this procedure, Land showed that a particular L, M, S quanta catch can appear any color.

James Clerk Maxwell invented colorimetry based on color matches. Also, he invented color photography based on three-color-separations with silver halide photographs. Both colorimetry matches and silver halide photography are molecular quanta catch events. It is easy to think that human vision uses those same mechanisms. But, color constancy disproves that hypothesis. Edwin Land's Retinex invention adopted the cone quanta catch as vision's initial receptor response but realized that vision's color appearance separations are the result of neural spatial comparisons, not receptors' cone quanta catch. Spatial comparisons are the basis of color constancy in complex scenes. Spatial comparisons form vision's L, M, S color separations.

These constancy observations led Land to propose the Retinex theory. The triplet of apparent lightnesses, not cone quanta catches, determines the color appearance. Constant LMS lightnesses generate constant colors. That hypothesis led to a study of color appearances in L, M, S bands of light. Do all red colors have the same triplet of

lightness appearances? Does a red color always look [light, dark, and dark] in L, M, S light? Does a green always look [dark, light, and dark]? Does color appearance always correlate with the triplet of L, M, S lightnesses?

The experiment is easy. Find narrowband red, green, and blue filters. Be sure that the filters exclude the other two-thirds of the spectrum. With a green filter, you should just see greens with different lightnesses. You should not see a mixture of green and yellow and blue hues. If you do, you need a filter with a narrower band of transmission.

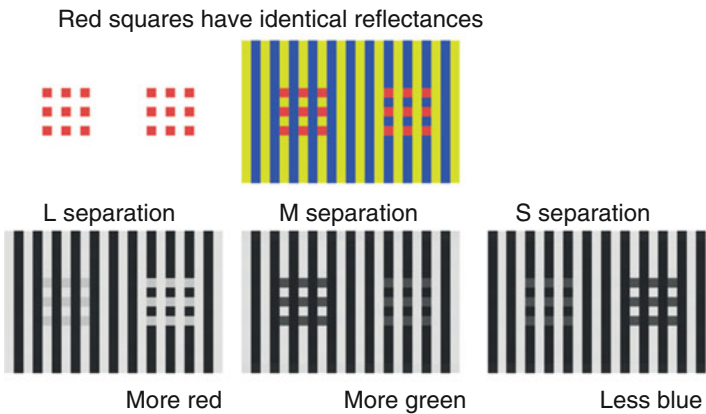
Identify a group of red objects. Look at them sequentially through the L, M, S filters. Look at them in different ambient illuminations. Look at them at different times of the day. Look at them in sunlight and shadows. Red colors are always (light, dark, and dark) in L, M, S light. The same dependence of the triplet of L, M, S lightnesses holds for all colors (Table 1). Lightness is the output of spatial image processing. It is the result of post-receptor spatial processing. That is why lightness does not correlate with cone quanta catch. However, color does correlate with three lightnesses in long-, middle-, and short-wave light.

Retinex Theory, Table 1 Correlation table of color appearances and the apparent lightnesses in L, M, S illumination

Color Appearance	Appearance in L- light	Appearance in M- light	Appearance in S- light
Red	light	dark	dark
Yellow	light	light	dark
Green	dark	light	dark
Cyan	dark	light	light
Blue	dark	dark	light
Magenta	light	dark	light
White	light	light	light
Black	dark	dark	dark

Retinex Theory, Fig. 5

(Top-left) Red color squares on white and (top-center) color assimilation experiments change the appearance of the red squares; (bottom) L, M, S separation images. Color appearances of red squares correlate with L, M, S lightnesses



Retinex theory predicts that the triplet of L, M, S lightnesses determines color. Colors are constant with changes in illumination because the triplet of lightnesses is nearly constant.

Land’s observation still stands: The triplet of lightnesses correlates with color. The observation is important because a variety of different phenomena can influence lightness, such as simultaneous contrast, the Cornsweet effect, and assimilation. Regardless of the cause of the lightness changes, when two identical physical objects look different, color appearances correlate with their L, M, S lightnesses ([7], Section E).

In the color assimilation display (Fig. 5), there are two sets of nine red squares that have the same reflectance and appear the same (top left).

However, if these red squares are surrounded by yellow and blue stripes, they look different (top center): The left red squares fall on top of the yellow stripes, and the right ones on the blue stripes. The left squares appear a purple red, while the right ones appear a yellow orange. In other words, the left squares appear more blue and the right ones more yellow.

In Fig. 5 (bottom-left), in the L separation, corresponding squares are lighter on the right of the separation; in the M separation, these patches are lighter; and in the S separation, they are darker on the right. Land’s Retinex predicts that whenever L and M separations are lighter and S separation is darker, the patch will appear more yellow (see right squares in assimilation

experiment (top-center)). Whenever S separation is lighter and L and M separations are darker, the patch will appear more blue (see left squares in assimilation experiment (top-center)). Colors correlate with L, M, S lightnesses ([7], pp. 221–281).

Retinex Image Processing

Land described that the fundamental challenge of color vision shifted to the ability to predict lightness, i.e., the spatial interactions in each color channel found in postreceptor neural processes. In 1967, Land and McCann proposed a computational model for calculating lightness from the array of all scene radiances [6]. The model compared each pixel with every other pixel in an image. The goal was to calculate the sensation of image segments that equaled what observers saw. In the past 50 years, there have been many implementations and variations of this process. They are called Retinex algorithms. It is curious that Land reserved the use of the term “Retinex” to describe three independent lightness channels. Today’s usage of the word includes a much wider range of computer algorithms that build calculated appearances out of arrays of radiances.

To calculate lightnesses in complex scenes, one must:

- Capture radiances from the entire scene.
- Convert scene radiances to cone and rod quanta catches.
- Calculate lightness for each color channel using all pixels in the scene.
- Compare calculated lightnesses with observer matches.

The Land and McCann model [6, 9, 11, (7, pp. 293–337)] used:

- Edge ratios of cone quanta catches (basic building block).
- Gradient threshold (found to be unnecessary in later studies).
- Multiplication of edge ratios (makes long-distance interactions).

- Reset to maxima (scaled the output).
(introduced dependence on scene content, e.g., simultaneous contrast)
- Average of many spatial comparisons.

The first computer implementation of the model in the early 1970s used an array of 20 by 24 pixels. McCann, McKee, and Taylor showed that long-, middle-, and short-wave computed lightnesses predicted observer matches of color Mondrians in color constancy experiments [8].

Since the late 1960s, computer imaging has shown remarkable advances. Digital images have replaced film in most of photography. Computer graphics has made image synthesis ubiquitous. Retinex image processing has grown with the advances in digital imaging [9–11]. In the early 1980s, Frankle and McCann introduced a multiresolution algorithm that allowed efficient comparison of all pixels in a 512 by 512 pixel image [12]. Multiresolution software programs (Zoom Retinex) calculate all scene lightnesses with $O(N)$ computational efficiency [9]. Sobol extended Frankle and McCann’s algorithm in the design of commercial digital cameras [13] introduced in 2003. Other applications used Retinex’s efficient spatial processing for color gamut-mapping [14].

Using different approaches, Jobson and Kotera have studied the NASA Retinex. Rizzi and colleagues have developed the Milan Retinex ([7], pp. 324–328).

The important feature of real complex scenes is that the illumination is rarely uniform. Shadows and multiple reflections increase the dynamic range of light coming to our eyes, and to cameras. The application of Retinex algorithms to high dynamic range (HDR) scenes has become a major topic of research and engineering applications ([7], pp. 283–375). The limits of HDR scene capture and reproduction are controlled by optics, namely, optical veiling glare. Camera glare limits the range of light on the sensor, just as intraocular glare limits the range of light on the retina. The scene content controls the distribution of glare range of light in images. Vision’s postreceptor neural processes compensate for veiling glare ([7],

pp. 89–171). That explains humans' high dynamic range of appearances from low-dynamic-range retinal images. The spatial mechanisms modeled by Retinex algorithms play a major role in compensating for glare and generating our range of color and lightness sensations [15, 16].

Over the years, many variations of spatial processing mimicking human vision have been called Retinex algorithms.

Signatures of Color Constancy

Studies of color constancy have documented its limits. Measurements of departures from perfect constancy provide a signature of the underlying postquantum-catch neural mechanisms. Land's double Mondrian experiments used spatially uniform illumination with different spectral compositions. In each spectral, illuminant observers chose matching Munsell chips very close to the actual Munsell papers in the Mondrian areas, indicating near-perfect constancy. Nevertheless, each spectral illuminant had different small systematic departures from perfect constancy across the entire Mondrian [8]. Similar small departures were observed in experiments with color adapting surrounds, again in uniform spatial illumination ([7], pp. 239–245). A pair of recent experiments measured color constancy using two identical 3D Mondrians made of blocks of wood painted with 1 of 11 matte paints. The experiments used different illuminations. The first used a diffusing tent that minimized shadows – the low-dynamic-range (LDR) scene. The second used spotlights that made sharp shadows – the high-dynamic-range (HDR) scene. Observers reported a marked loss of color constancy in 3D Mondrians. LDR illumination caused a substantial loss of constancy in some local regions; HDR illumination caused a much larger loss in other regions [17]. The appearance of a surface in 3D blocks no longer shows universal correlation with the paint's surface reflectance when the illumination introduces quanta-catch edges. These and other experiments support Retinex's spatial comparison mechanism [18]. Five different studies of the signatures of color appearance in real scenes demonstrate the

important role of spatial comparisons in vision [19].

Summary

Edwin Land coined the term Retinex in 1964 as the name of human color channels that build up appearance from spatial comparisons. Over the years, others have redefined the term to include many different spatial algorithms that attempt to mimic human spatial color processing. Land's Retinex was built to model color constancy. Over the years, quantitative measurements of the limits of constancy have provided signatures of the underlying visual mechanism. These signatures support neural spatial comparisons within color channels.

Cross-References

- [Color Constancy](#)
- [Glare](#)

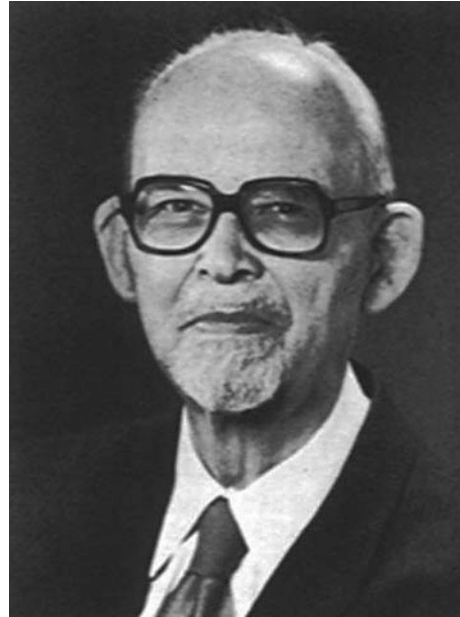
References

1. McCann, J.J., Benton, J., McKee, S.: Red/white projections and rod/long-wave cone color: an annotated bibliography. *J. Electron. Imaging*, **13**, 8–14 (2004)
2. Land, E.H.: The retinex. *Am. Sci.* **52**, 247–264 (1964)
3. Land, E.H. McCann, J.J.: Method and system for reproduction based on significant visual boundaries of original subject, U.S. Patent 3,553,360, June 5, 1971
4. Land, E.H.: The retinex theory of colour vision. *Proc. R. Inst. Gr. Brit.* **47**, 23–58 (1974)
5. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**, 108–128 (1977)
6. Land, E.H., McCann, J.J.: Lightness and retinex theory. *J. Opt. Soc. Am.* **61**, 1–11 (1971)
7. McCann, J., Rizzi, A.: *The Art and Science of HDR Imaging*, pp. 221–375. Wiley, Chichester (2012)
8. McCann, J.J., McKee, S., Taylor, T.: Quantitative studies in retinex theory, a comparison between theoretical predictions and observer responses to color Mondrian experiments. *Vis. Res.* **16**, 445–458 (1976)
9. McCann J.J.: Lessons learned from Mondrians applied to real images and color gamuts. In: *Proceedings of IS&T/SID Color Imaging Conference*, vol. 7, pp. 1–8. Scottsdale (1999)

10. McCann, J.J.: Simultaneous contrast and color constancy: signatures of human image processing. Chapter 6. In: *Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives*, Vancouver Studies in Cognitive Studies: Davis, A., Davis, S. (eds.) pp. 87–101. Oxford University Press, Vancouver (2000)
11. McCann, J.J.: Capturing a black cat in shade: past and present of retinex color appearance models. *J. Electron. Imaging*. **13**, 36–47 (2004)
12. Frankle, J., McCann, J.: Method and apparatus of lightness imaging. US Patent 4,384,336 (1983)
13. Sobol, R.: Improving the retinex algorithm for rendering wide dynamic range photographs. *J. Electron. Imaging*. **13**, 65–74 (2004)
14. McCann, J.J.: A spatial color-gamut calculation to optimize color appearance. In: MacDonald, L., Luo, R. (eds.) *Colour Image Science: Exploiting Digital Media*, pp. 213–233. Wiley, Chichester (2002)
15. McCann, J.J., Vonikakis, V., Rizzi, A.: HDR Scene Capture and Appearance. SPIE Spotlight Tutorial, Bellingham (2018). https://spie.org/Publications/Book/2315540?&origin_id=x109925&SSO=1
16. McCann, J.J., Vonikakis, V.: Calculating retinal contrast from scene content: a program. *Front. Psychol.*, 17 Jan (2018). <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02079/full>
17. McCann, J.J., Parraman, C., Rizzi, A.: Reflectance, illumination, and appearance in color constancy. *Front. Psychol.* **24**, 00005 (2014). <https://doi.org/10.3389/fpsyg.2014.00005>
18. McCann, J.J.: Retinex at 50: color theory and spatial algorithms, a review. *J. Electron. Imaging*. **26**(3), 031204 (2017). <https://doi.org/10.1117/1.JEI.26.3.031204>
19. McCann, J.J.: Limits of Color Constancy: comparison of signatures of chromatic adaptation and spatial comparisons. In: *Proceedings of Electronic Imaging: Color Imaging XXIV, IS&T*, San Francisco (2019). <https://doi.org/10.2352/ISSN.2470-1173.2019.14.COLOR-085>

Richter, Manfred

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Richter was born on August 7, 1905, in Dresden, Germany, where he studied technical physics under Robert Luther at the Technical University from 1924 to 1933. The subject of his doctoral dissertation was Goethe's *Farbenlehre* as related to scientific problems. [1] In 1927, as an assistant in the department of color research of the German Institute of Textile Research and following Helmholtz' assistant Arthur König, he developed an international bibliography of publications in color science, an effort he continued until the mid-1950s [2]. In 1934, he began work in the laboratories of the lamp manufacturer OSRAM in Berlin. In 1943, he transitioned to the *Materialprüfungsanstalt* (Office for testing of materials), later named *Bundesanstalt für Materialprüfung* (BAM), where he remained until 1962 and where he organized a color research laboratory. In 1941, he was asked by *Deutsches Institut*

Retinogeniculate Pathways

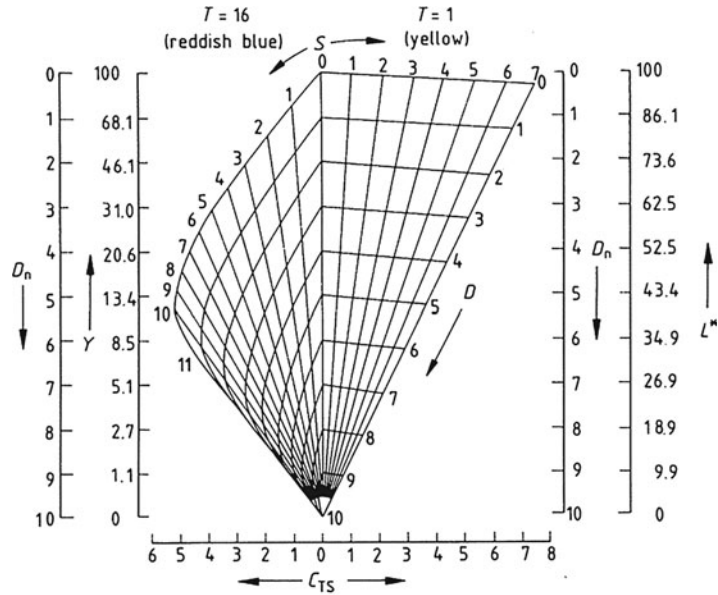
► Magno-, Parvo-, and Koniocellular Pathways

Retinogeniculocortical Pathways

► Magno-, Parvo-, and Koniocellular Pathways

Richter, Manfred,

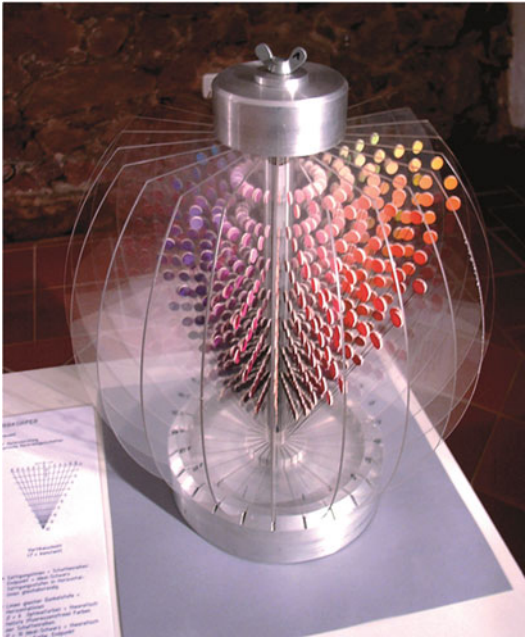
Fig. 1 Schematic representation of the samples of hues T1 and T16 in the S, D diagram. (Ref. [3])



für Normung (DIN, German institution for industrial standards) to develop a standard color system and atlas, an effort that kept him occupied for an extended period of time. It is known as DIN6164 today. He was also a professor at the *Institut für Lichttechnik* (Institute for lighting technology) of the Berlin Technical University. In 1949, he was a founding member of the *Fachnormenausschuss Farbe* (FNF, Color standards committee). He was a leading member of *Deutsche farbwissenschaftliche Gesellschaft* (DfwG, German society for color science, founded in 1974) and a founding member of the journal *Die Farbe* (1951–2003). He was a member of the directorial board of the International Association of Colour (AIC) and active in several research committees of the International Association of Illumination (CIE). His passing on April 20, 1990, was the result of a traffic accident.

Major Accomplishments/Contributions

DIN6164: Richter’s plan for a standard object color system and atlas was that it needed to be based on well-supported colorimetric data and the latest insights into color order. A starting point was Ostwald’s color atlas and the Luther-Nyberg optimal object color solid. The system was to be perceptually uniform. As perceptual parameters he



Richter, Manfred, Fig. 2 3D representation of the samples of the DIN6164 system. © Eckhard Bendin

selected hue (T), saturation (S), and degree of darkness (D). Lacking a satisfactory colorimetric model of hue scaling he proceeded to experimentally determine a constant saturation contour in the CIE chromaticity diagram separated into 24 perceptually equal hue differences. He then scaled saturation

from the neutral point to the spectral limit into up to 16 levels. The darkness scale D is based on logarithmic scaling of the relative brightness value scale proposed in 1928 by S. Rösch. It has a value of 0 for white and 10 for black. A schematic cross section of the system is shown in Fig. 1.

Atlases were published in 1960/62 with matte samples and 1978/83 with glossy samples. Figure 2 shows a 3D model of the system. As its name indicates, DIN6164 is a German industrial standard system.

Publications: In 1940 Richter published a book on color science, with the cooperation of I. Schmidt and A. Dresler, that presented the subject in at that time likely the most comprehensive and detailed form. [4] Given the Second World War it was never translated into English. In 1976, he published *Einführung in die Farbmatrik* (Introduction to color metrics) [5].

References

1. Richter, M.: Das Schrifttum über Goethes Farbenlehre, mit besonderer Berücksichtigung der naturwissenschaftlichen Probleme. Pfau, Berlin (1938)
2. Richter, M.: Internationale Bibliographie der Farbenlehre und ihrer Grenzgebiete, 2 vols, Musterschmidt, Göttingen (1940–1949 and 1950–1954)
3. Richter, M., Witt, K.: The story of the DIN color system. *Color Research and Application*. **11**, 138–148 (1986)
4. Richter, M. (with collaboration of I. Schmidt and A. Dresler). *Grundriss der Farbenlehre der Gegenwart* (Basics of the color science of the present). Steinkopff, Dresden (1940)
5. Richter, M.: *Einführung in die Farbmatrik*. de Gruyter, Berlin (1976)

Road Lighting

John D. Bullough

Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

Synonyms

Highway lighting; Public lighting; Roadway lighting; Street lighting

Definition

Road lighting is the application of illumination systems along roadways, primarily for the purpose of improving safety by increasing visibility of roadside hazards and by reducing the effects of glare from other light sources in the visual environment, such as vehicle headlamps.

Introduction

Road lighting systems are an important part of the highway safety infrastructure across the world. Two important purposes of road lighting are to allow drivers to see the roadway further ahead than their own vehicle's headlamps allow and to reduce glare from other vehicles' headlamps [1]. Indeed, road lighting is generally associated with reduced nighttime crash rates [2, 3], and the presence of illumination that results in increased luminances of potential road hazards can reduce the impact veiling luminances produced by bright sources of light in the driver's field of view. In the present article, existing practices and standards for road lighting are described, as well as the light source and luminaire technologies used to provide road lighting. For further details on road lighting and interactions with other visual information systems, consult references [4, 5].

Practices and Standards

CIE 115

In most of the world, the primary road lighting standard is the Commission Internationale de l'Éclairage (CIE) *Lighting of Roads for Motor and Pedestrian Traffic*, CIE 115 [6]. In CIE 115, three main lighting classes are specified: for motorized traffic (M), for conflict areas between vehicles or between vehicles and pedestrians (C), and for pedestrian traffic (P). Each main class is subdivided into numerical classes based upon factors such as traffic volume and speed, road access control, ambient environment, and the mix of motorized and pedestrian traffic. Lower-numbered classes refer to more complex road situations requiring higher light levels.

Specifications for M classes M1 through M6 are given in average road surface luminance and range from 2 to 0.3 cd/m^2 . For C classes C0 through C5, illuminance specifications range from 50 to 7.5 lx. For P classes P1 through P6, illuminance specifications range from 15 to 2 lx. CIE 115 specifications also include limits for characteristics such as uniformity, glare, and, when facial recognition of pedestrians is important, vertical illuminance [6].

CIE 115 also provides a method for the specification of adaptive lighting to account for the possibility of reduced vehicle or pedestrian traffic at certain times throughout the night [6]. Light levels can be reduced to different classes, resulting in decreased energy use and light pollution.

Road lighting calculations are often performed using software that accepts a photometric file containing tabulated luminous intensity data for a luminaire. For a set of road geometric characteristics, pole locations and heights, the software calculates the illuminances or luminances, uniformities, and veiling luminances. Such calculations should account for lamp output reductions and dirt accumulation over time. The light output from a road lighting system could be 20–30% lower after several years than when it was new and clean.

RP-8

In North America, the primary standard for most road lighting is the Illuminating Engineering Society (IES) *Recommended Practice for Design and Maintenance of Roadway and Parking Facility Lighting*, RP-8 [1]. RP-8 serves as the basis for continuous road lighting system by the American Association of State Highway and Transportation Officials [AASHTO [7]], which provides warrants for road lighting based on traffic volume and crash frequency at night. RP-8 includes different specification methods for different facility types in terms of the luminance of the roadway surface. There are also requirements for uniformity of luminance and limits to control for glare in terms of veiling luminance.

The target light levels depend upon the road type (i.e., local, collector, major, or freeway) and the pedestrian conflict level. For example, for a

local road with a medium pedestrian conflict level, the average luminance should be at least 0.5 cd/m^2 , the minimum luminance no less than 1/6 the average, and the veiling luminance should not exceed 40% of the average road surface luminance [1]. For a major roadway with the same pedestrian conflict level, the average luminance should be at least 0.9 cd/m^2 , the minimum luminance should be no less than 1/3 the average, and the veiling luminance should not exceed 30% of the average road surface luminance [1]. Calculations for RP-8 road lighting criteria are performed similarly as for CIE 115 criteria.

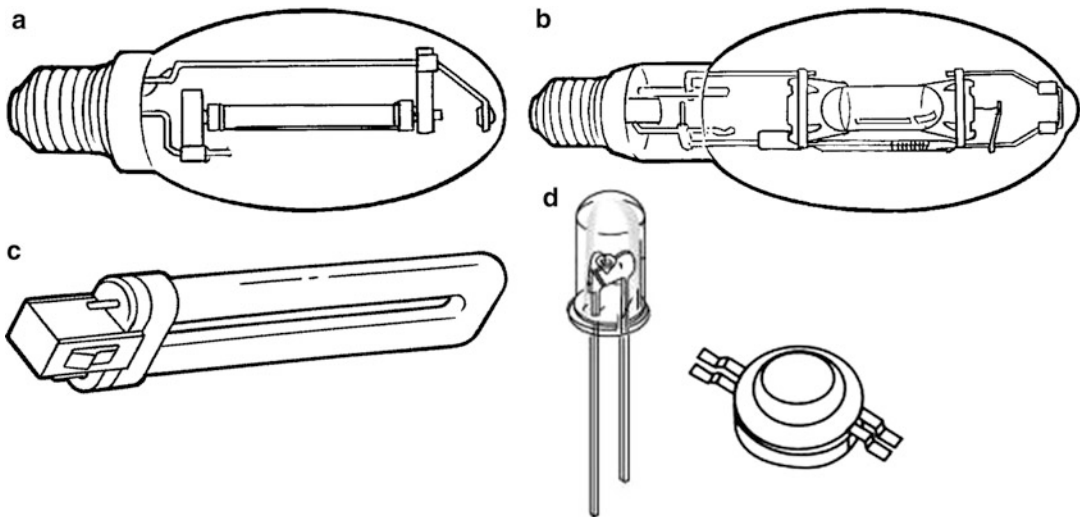
RP-8 criteria are for continuous lighting along roads excluding intersections and interchanges. Intersection lighting criteria are given in terms of illuminances on the road surface, based on the types of intersecting roads. Like CIE 115, RP-8 permits the implementation of adaptive lighting when pedestrian (or vehicular) use changes throughout the night.

Light Sources

Road lighting systems can use a wide variety of light sources. At present (as of 2019) most road lighting installations use discharge sources (high- and low-pressure sodium, metal halide, fluorescent, and induction lamps). Increasingly, however, solid-state (light-emitting diodes) sources are being used for road lighting. Several common light sources for road illumination are illustrated in Fig. 1.

High-Pressure Sodium

High-pressure sodium (HPS) lamps emit light when a current is applied to sodium vapor. The inner envelope of an HPS lamp contains sodium, and the outer glass bulb absorbs ultraviolet (UV) energy and stabilizes the arc tube temperature. HPS lamp efficacies are high, and they have long operating lives and very high lumen maintenance. HPS lamp illumination is yellowish in color. “Whiter” HPS lamps can be developed by increasing the sodium vapor pressure or operating HPS lamps at high frequencies. HPS lamps are the most common sources used in road lighting. For



Road Lighting, Fig. 1 (a) High-pressure sodium (HPS) lamp (Courtesy of the Lighting Research Center). (b) Metal halide (MH) lamp (Courtesy of the Lighting Research Center). (c) Fluorescent lamp (Courtesy of the

Lighting Research Center). (d) Two light-emitting diode (LED) packages (Courtesy of the Lighting Research Center)

additional details about the performance of HPS lamps, consult the entry ► [“High- and Low-Pressure Sodium Lamp”](#) in this Encyclopedia.

Metal Halide

Metal halide (MH) lamps have an arc tube containing various metallic halide compounds in addition to mercury, which emit light across the visible spectrum. MH lamps have relatively high efficacy and are available in a range of white colors ranging from “warm” to “cool” in appearance. Lamp life is typically not as high as HPS lamps but has been improving substantially in recent years. Lumen maintenance values range from fair to good. MH lamps with ceramic arc tubes and improved starting gear are available, which have much-improved life and lumen maintenance. For additional details about the performance of MH lamps, consult the entry ► [“Metal Halide Lamp”](#) in this Encyclopedia.

Low-Pressure Sodium

Like HPS, low-pressure sodium (LPS) lamps use sodium vapor but at a much lower vapor pressure. LPS lamps are linear in shape. They produce nearly monochromatic yellow light near 589 nm. They have relatively long operating lives, excellent lumen maintenance, and very high luminous

efficacy, although color rendering is essentially nonexistent with these lamps. LPS lamps are sometimes used near astronomical observatories because it is fairly easy to filter out the wavelengths emitted this source. LPS is more common in Europe than in North America for road lighting. For additional details about the performance of LPS lamps, consult the entry ► [“High- and Low-Pressure Sodium Lamp”](#) in this Encyclopedia.

Fluorescent and Induction

Fluorescent lamps are low-pressure gas discharge sources. Light is produced by fluorescent phosphors activated by ultraviolet energy from a mercury arc. Fluorescent lamps have reasonably long operating lives, limited primarily by the life of the lamp end electrodes that generate the electrical discharge. Fluorescent lamps require ballasts to provide the starting and operating voltages and currents needed for proper lamp operation. Fluorescent lamp color of fluorescent lamps is determined by the phosphors used to coat the envelope. Fluorescent lamps have relatively higher luminous efficacies and lumen maintenance values.

A special fluorescent lamp type, the induction lamp, has found some success for road lighting. Induction lamps have somewhat more compact shapes than tubular fluorescent lamps because

they use a magnetic induction coil to provide the current that stimulates the mercury vapor inside the lamp. Typical operating lives are double or triple those of fluorescent lamps, with luminous efficacy comparable in value. For additional details about the performance of fluorescent and induction lamps, consult the entries ► [“Tubular and Compact Fluorescent Lamp”](#) and ► [“Induction Lamp”](#) in this Encyclopedia.

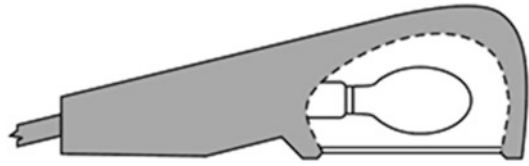
Light-Emitting Diodes

Light-emitting diodes (LEDs) are solid-state semiconductor junction devices that emit light when a current is passed through the junction. White light can be created by mixing light from red, green, and blue LEDs, or by using blue LEDs with phosphors that convert some of the blue light into yellow light, resulting in a mixture that is perceived as white. Newer LED packages exist in addition to 5-mm diameter epoxy capsule LEDs used for indicator lights. Some packages contain metal heat sinks to dissipate internal temperatures, which can reduce LED light output and shorten operating life. LED lighting technology is increasing in luminous efficacy and LEDs also have very long operating lives with gradual reductions in light output over time. For additional details about the performance of LEDs, consult the entry ► [“Light-Emitting Diode, LED”](#) in this Encyclopedia.

Luminaires

Road lighting systems generally consist of pole-mounted luminaires, usually with individual photocell or timer control. The most common luminaire type used for road lighting is the so-called cobrahead luminaire (Fig. 2), named for its distinctive shape. Figure 3 shows typical road luminaire construction. Luminaires are designed for a particular type of light source such as discharge lamps including HPS or MH [8], and there is increasing interest in induction and especially LED luminaire systems [9, 10].

Luminaires for road lighting can be classified in several ways. A cutoff classification system [1], used by many North American road lighting specifiers, classifies luminaires according to their



Road Lighting, Fig. 2 Cobrahead luminaire for road lighting (Courtesy of the Lighting Research Center)

luminous intensity distributions in different angular regions (relative to nadir, defined as 0° directly below the luminaire) where light output could contribute to glare or light pollution. Table 1 summarizes the luminous intensity limits (in cd) for different cutoff classifications, relative to the light output (in lm) of the lamp inside the luminaire. The IES has adopted a system to classify the distribution of light from an outdoor luminaire that simply uses the amount of light output (in lm) emitted within various angular regions to serve as the basis for classification [12].

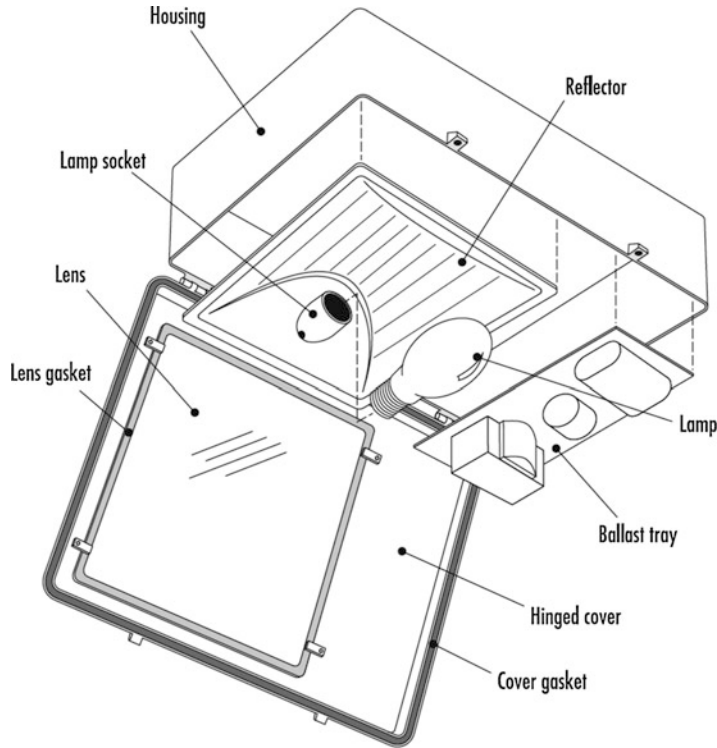
As mentioned previously, the most common control system for roadway lighting is a photocell mounted to an individual luminaire, which will switch the luminaire on and off at a specified ambient light level. Some systems will use centralized control via a single photocell or time clocks. Technological innovations are making centralized control systems attractive, because these systems can also monitor performance of individual luminaires and alert the system operator when a lamp or ballast failure has occurred or is close to occurring [13]. They can also be useful in controlling adaptive road lighting systems.

Color: Mesopic Vision

The color of the light source used for road lighting can influence visual perception of drivers and pedestrians. Although photometric specifications for road lighting are based on the photopic luminous efficiency function, representing the spectral sensitivity of cone visual photoreceptors to light at daytime light levels, visual spectral sensitivity differs at low levels commonly experienced under road lighting at night. At very low light levels, rod photoreceptors, whose sensitivity is represented by the scotopic luminous efficiency

Road Lighting,

Fig. 3 Components within a typical road lighting luminaire (Courtesy of the Lighting Research Center)



Road Lighting, Table 1 Light output limits for luminaire cutoff classifications [11], given in luminous intensity values (in cd) as a percentage of the luminaire’s lamp lumens (in lm)

Cutoff type	Light between 80° and 90°	Light above 90°
Full cutoff	100 cd per 1,000 lm	None
Cutoff	100 cd per 1,000 lm	25 cd per 1,000 lm
Semicutoff	200 cd per 1,000 lm	50 cd per 1,000 lm
Noncutoff	No limitation	No limitation

function, dominate vision. At many road lighting levels, both rods and cones contribute to vision and luminous efficiency can be approximated by a combination of photopic and scotopic efficiency [14].

As a consequence, light sources with relatively greater power in the short-wavelength portion of the visible spectrum can be more effective for vision at these so-called mesopic, nighttime light levels. Lamp spectra can be characterized by their scotopic/photopic (S/P) ratio [15], which

Road Lighting, Table 2 Scotopic/photopic (S/P) ratios of common light sources [14, 15]

Light source	S/P ratio
HPS 250 W clear	0.63
MH 400 W clear	1.51
LPS	0.25
Fluorescent/induction (“cool white”)	1.48
LED (4,300 K)	2.04

indicates a light source’s ability to stimulate the rods for an equivalent amount of cone stimulation. Table 2 lists S/P ratios for several road lighting sources. The CIE has developed a unified system of photometry [15] bridging the photopic and scotopic systems, as part of CIE Publication 191 [15], to quantify mesopic luminance. Above a luminance of 5 cd/m², the unified luminance and the photopic luminance are equivalent, and below a luminance of 0.005 cd/m², the unified luminance and the scotopic luminance are equivalent. The CIE 191 system uses the photopic luminance and the S/P ratio of the light source under investigation to estimate the unified luminance.

Cross-References

- [Automotive Front Lighting System \(Status 2020\)](#)
- [Glare](#)
- [High- and Low-Pressure Sodium Lamp](#)
- [Induction Lamp](#)
- [Light Pollution](#)
- [Light-Emitting Diode, LED](#)
- [Luminaires](#)
- [Metal Halide Lamp](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Illuminating Engineering Society: Recommended Practice for Design and Maintenance of Roadway and Parking Facility Lighting, RP-8. Illuminating Engineering Society, New York (2018)
2. Commission Internationale de l'Éclairage: Road Lighting as an Accident Countermeasure, CIE 93. Commission Internationale de l'Éclairage, Vienna (1992)
3. Bullough, J.D., Rea, M.S.: Intelligent control of roadway lighting to optimize safety benefits per overall costs. Paper presented at the 14th Institute of Electrical and Electronics Engineers conference on intelligent transportation systems, George Washington University, Washington, 5–7 October 2011
4. Boyce, P.R.: Lighting for Driving. CRC Press, New York (2009)
5. Bullough, J.D.: Roadway transportation lighting. In: Kutz, M. (ed.) Handbook of Transportation Engineering, vol. II, 2nd edn. McGraw-Hill, New York (2011)
6. Commission Internationale de l'Éclairage: Lighting of Roads for Motor and Pedestrian Traffic, CIE 115. Commission Internationale de l'Éclairage, Vienna (2010)
7. American Association of State Highway and Transportation Officials: Roadway Lighting Design Guide, GL-7. American Association of State Highway and Transportation Officials, Washington, DC (2018)
8. McColgan, M., Van Derlofske, J., Bullough, J.D., Vasconez, S.: Specifier Reports: Parking Lot and Area Luminaires. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2004)
9. Radetsky, L.: Specifier Reports: Streetlights for Collector Roads. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2010)
10. Radetsky, L.: Specifier Reports: Streetlights for Local Roads. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2011)
11. Rea, M.S. (ed.): Illuminating Engineering Society Lighting Handbook: Reference and Application, 9th edn. Illuminating Engineering Society, New York (2000)
12. Illuminating Engineering Society: Luminaire Classification System for Outdoor Luminaires, TM-15. Illuminating Engineering Society, New York (2011)
13. Bullough, J.D.: Lighting Answers: Dynamic Outdoor Lighting. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2010)
14. Rea, M.S., Bullough, J.D., Bierman, A., Freyssinier-Nova, J.P.: A proposed unified system of photometry. Light. Res. Technol. **36**(2), 85–111 (2004)
15. Commission Internationale de l'Éclairage: Recommended System for Mesopic Photometry Based on Visual Performance, CIE 191. Commission Internationale de l'Éclairage, Vienna (2010)

Roadway Lighting

- [Road Lighting](#)

Runge, Philipp Otto

Rolf G. Kuehni
Charlotte, NC, USA



P. O. Runge, Self-portrait, ca. 1804

Biography

Runge was born on July 23, 1777, the 9th of 11 children of a tradesman and cargo ship owner and his wife in Wolgast, Pomerania, on the Baltic Sea, then under Swedish rule. As a child he was frequently ill with tuberculosis, being often educated at home. In 1795 he began a commercial apprenticeship at his older brother Daniel's firm in Hamburg. In 1799 Daniel supported Otto financially to begin the study of painting at the Copenhagen Academy. In 1801 Otto moved to Dresden to continue his studies, where among others he met his future wife Pauline Bassenge and the painter Caspar David Friedrich. He also began to study extensively the writings of the seventeenth-century mystic Jakob Boehme. In 1803, on a visit to Weimar, Runge unexpectedly met Johann Wolfgang von Goethe, and the two formed a friendship based on their common interests in color and art. After marrying Pauline in 1804, they moved to Hamburg. But due to imminent war dangers (Napoleonic siege of Hamburg), they relocated in 1805 to his parental home in Wolgast where they remained until 1807, then returned to Hamburg. Together they had four children, the last born on the day after Runge's premature death in 1810. In March of 1810 Runge became ill with tuberculosis again to which he succumbed on December 2 of the year. Runge was of a mystical, deeply Christian turn of mind, and in his artistic work he tried to express notions of the harmony of the universe through symbolism of color, form, and numbers. He considered blue, yellow, and red to be symbolic of the Christian trinity, equating blue with God and the night; red with morning, evening, and Jesus; and yellow with the Holy Spirit (Runge 1841, I, p. 17 [1]). During his short life he became one of the most important German painters of the Romantic period.

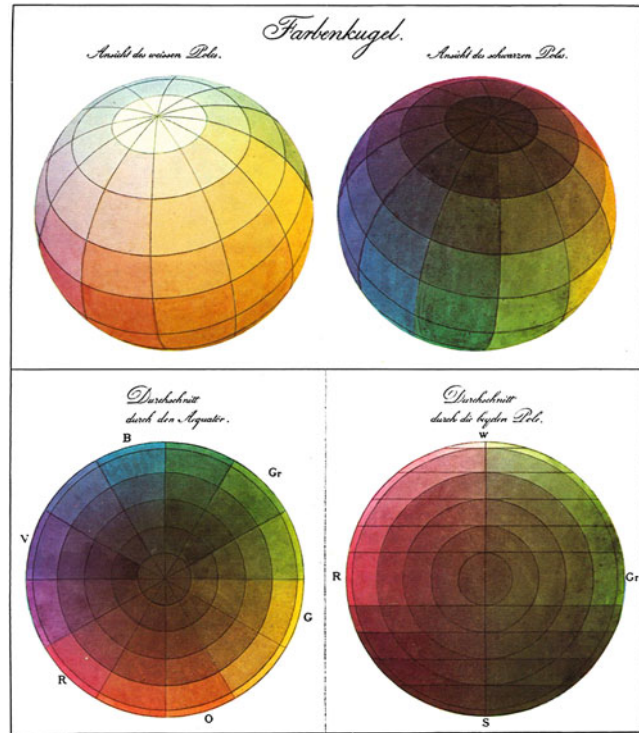
Major Accomplishments/Contributions

Runge's interest in color was the natural result of his work as a painter and of having an

enquiring mind. Among his accepted tenets was that "as is known, there are only three colors, yellow, red, and blue." (letter to Goethe of July 3, 1806) [2]. His goal was to establish the complete range of colors resulting from mixture of the three primaries, among themselves and together with white and black. In the same lengthy letter, Runge discussed in some detail his views on color order and included a sketch of a mixture circle, with the three primary colors forming an equilateral triangle and, together with their pair-wise mixtures, a hexagon. He arrived at the concept of the color sphere sometime in 1807, as indicated in his letter to Goethe of November 21 of that year, by expanding the hue circle into a sphere, with white and black forming the two poles [3]. A color mixture solid of a double-triangular pyramid had been proposed by Tobias Mayer in 1758 and implemented in form of a triangular pyramid by J. H. Lambert, facts known to Runge. His expansion of that solid into a sphere appears to have had an idealistic basis rather than one of logical necessity. In 1808 he hoped to provide scientific support for the sphere form with disk color mixture experiments. Encouraged by Goethe and other friends, he wrote in 1808 a manuscript describing the color sphere, published in Hamburg early in 1810 as *Farben-Kugel* (Color sphere) by his friend Friederich Perthes [4]. An included hand-colored plate shows two different views of the surface of the sphere as well as horizontal and vertical slices demonstrating the organization of its interior (Fig. 1).

In addition to a description of the color sphere, it contains an illustrated essay on rules of color harmony by Runge and one on color in nature written by Runge's friend Henrik Steffens. Runge's premature death limited the impact of this work. Goethe, who had read the manuscript before publication, mentioned it optimistically in his *Materialien zur Geschichte der Farbenlehre* of 1810 as "successfully concluding this kind of effort" [5]. More detailed similar systems were published soon thereafter

Runge, Philipp Otto,
Fig. 1 Hand-illuminated
 illustration from P. O.
 Runge, *Farben-Kugel*,
 Ref. [4]. Top: Image of
 sphere top down (*top left*)
 and bottom up (*top right*).
 Horizontal cross section
 through middle of the
 sphere (*bottom left*) and
 vertical cross section
 (*bottom right*)



by Gaspard Grégoire in France and by Mathias Klotz in Germany [6, 7].

References

1. Runge, P.O.: *Hinterlassene Schriften*, 2 vols. Hamburg, Perthes 1840/41
2. Letter by Runge to J. W. von Goethe dated July 3, 1806, in Maltzahn, H.: *Philipp Otto Runge's Briefwechsel mit Goethe*, Weimar: Verlag der Goethe-Gesellschaft This letter was included as an appendix in Goethe's *Farbenlehre* of 1810 (1940)
3. Letter by Runge to J. W. von Goethe dated Nov. 21, 1807, in Maltzahn, H., op. cit
4. Runge, P.O.: *Die Farben-Kugel, oder Construction des Verhaeltnisses aller Farben zueinander*, Perthes, Hamburg (1810) (English text translation available on the website www.iscc.org)
5. Goethe, J.W.: *Materialien zur Geschichte der Farbenlehre*. Cotta, Tübingen (1810). Entry on J. H. Lambert
6. Grégoire, G.: *Théorie des couleurs, contenant explication de la table des couleurs*. Brunot-Labbe, Paris (1815)
7. Klotz, M.: *Gründliche Farbenlehre*. Lindauer, München (1816)

Rushton, William A. H.

Stephen Westland

Colour Science and Technology, University of Leeds, Leeds, UK

William Albert Hugh Rushton was a British physiologist who made important contributions to our understanding of color vision and perception. He is perhaps best known now for his development of the principle of univariance.

Rushton was born in London on 8 December 1901. He entered Cambridge University as a medical student in 1921 and obtained a degree in physiology in 1925. He received a PhD degree in 1928 working under Prof. E.D. Adrian for his research investigating the flow of current in and around nerves to determine the portion of the current responsible for excitation. As a result of this work he won the Stokes Studentship at Pembroke College in 1929. He spent 2 years at the Johnson

Foundation in Philadelphia before returning to a Research Fellowship at Cambridge University. He went to University College Hospital in 1931 to study clinical medicine before obtaining a Lectureship at Cambridge University in 1935 after which he produced a prodigious body of work; he published 37 papers on nerve with 4 colleagues over 25 years and 147 papers on vision with 27 colleagues over a period of 30 years. Upon retirement he spent some time at Florida State University as a Distinguished Research Professor. He was awarded the Royal Medal of the Royal Society in 1970 and continued to publish until he died in June 1980.

Nervous Excitation

Rushton laid the groundwork for the establishment of the modern theory of nervous excitation and propagation by his quantitative analysis of the temporal and spatial factors involved in electrical excitation [1, 2].

Retinal Densitometry

He developed techniques to measure the visual pigments *in vivo*. And made important contributions regarding the distribution of pigments in the retina, the action spectrum of bleaching, and the spectral characteristics of bleaching [3–5].

Cone Pigments

By 1955, Rushton had evidence of visual pigments at the fovea in normal and color-blind subjects. He obtained evidence of two pigments in the medium-long wavelength spectrum in color normals and demonstrated that protonopes and deuteranopes were each lacking one of these pigments [6, 7].

Analytical Anomaloscope

In the early 1960s, Rushton developed an analytic anomaloscope to study color vision. Using this

instrument, observers would match a pure spectral light with an additive mixture of red and green in variable proportions [8].

Principle of Univariance

Rushton published two papers in *Scientific American* and gave several notable didactic lectures. In a review lecture at the Physiological Society (London), he described the principle of univariance that: “the output of a receptor depends upon its quantum catch, but not upon which quanta are caught.” Many of the laws of color mixing are a direct consequence of this principle. He also reintroduced the cone pigment triangle, first put forward by Maxwell, as an alternative to the usual representation of color space developed by the CIE [9].

References

1. Rushton, W.A.H.: The time factor in electrical excitation. *Biol. Rev.* **10**, 1–17 (1935)
2. Rushton, W.A.H.: A theory of excitation. *J. Physiol. Lond.* **84**, 42P (1935)
3. Campbell, F.W., Rushton, W.A.H.: The measurement of rhodopsin in the human eye. *J. Physiol. Lond.* **126**, 36P–37P (1954)
4. Campbell, F.W., Rushton, W.A.H.: Measurement of scotopic pigment in the living human eye. *J. Physiol. Lond.* **130**, 131–147 (1955)
5. Rushton, W.A.H.: The difference spectrum and the photosensitivity of rhodopsin in the living human eye. *J. Physiol. Lond.* **134**, 11–29 (1956)
6. Rushton, W.A.H.: A cone pigment in the protonope. *J. Physiol. Lond.* **168**, 345–359 (1963)
7. Rushton, W.A.H.: A foveal pigment in the deuteranope. *J. Physiol. Lond.* **176**, 24–37 (1965)
8. Baker, H.D., Rushton, W.A.H.: An analytical anomaloscope. *J. Physiol. Lond.* **168**, 31P–33P (1963)
9. Rushton, W.A.H.: Pigments and signals in colour vision (Review Lecture). *J. Physiol. Lond.* **220**, 1P–31P (1972)

Further Reading

Barlow, H.B.: William Rushton. 8 December 1901–21 June 1980. *Biograph. Mem. Fellows R. Soc.* **32**, 422–426 (1986)

RYB Color

Zena O'Connor
Design Research Associates, Sydney, NSW,
Australia

Synonyms

Designer's color model; Painter's color model;
RYB triad; Traditional color model

Definition

A cornerstone component of traditional color theory, the RYB conceptual color model underpins the notion that the creation of an exhaustive gamut of color nuances occurs via intermixture of red, yellow, and blue pigments, especially when applied in conjunction with white and black pigment color. In the literature relating to traditional color theory and RYB color, red, yellow, and blue are often referred to as primary colors and represent exemplar hues rather than specific hues that are more pure, unique, or proprietary variants of these hues. Under traditional color theory, the RYB color model is accompanied by a range of hierarchical color constructs and color combination constructs.

Categories of Color

Green-Armytage identified a number of color categories, and each of these is underpinned by specific and distinct ontological and epistemological assumptions and traditions [1]. These categories include conventional color, substance color, formula color, spectral profile color, psychophysical color, inherent color, and perceived color. In relation to these categories of color, RYB color falls under the specific category of substance color.

Traditional Color Theory

Traditional color theory is just one of various branches of color theory, and each of these branches tends to relate to a specific category of color, as detailed above [2]. Gage notes that a number of factors distinguish traditional color theories and sets these aside from other color theories: (a) an ontological focus on pigment color exploration and application in art, design, and architecture; (b) consideration of an exemplar set of primary pigment colors from which a large gamut of color nuances can be created; (c) a focus on creating and applying color harmony; and (d) the provision of color combination guidelines for color exploration and application [3].

Hierarchical Color Classifications

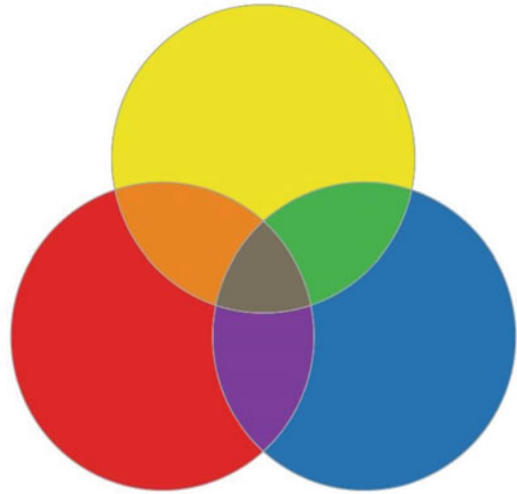
Under traditional color theory and the RYB color model, hierarchical color classifications are evident. The first of these classifications is referred to in the literature as “primary,” “principal,” “simple,” or “basic” colors, and in respect to RYB color these are red, yellow, and blue. The next level is referred to as “secondary” or “compound” colors, representing intermixture of primary colors, and under traditional color theory and RYB color these are orange, green, and purple. A third hierarchical level is referred to as “tertiary” color, representing an intermixture of red, yellow, and blue in varying proportions [2–19].

In respect to the RYB color classification of “tertiary” color, Ruskin muddled this term and contradicted the common understanding of the term by suggesting that “tertiary” colors were those that lay *between* “secondary” colors as this enabled him to avoid what he termed *tiresome* color names, an example of misleading definition that continues today [2, 20].

Origins and Development of RYB Color

Aside from commentary offered by ancient Greek philosophers, Tobias Mayer is credited with

RYB Color, Fig. 1 RYB color model (Illustration by Zena O'Connor and Geoffrey Gifford 2021)



conceptualizing and documenting the first RYB color model base [4–6].

Multiple traditional color theorists since Mayer presented color models based on the RYB triad supported by detailed rationale and discussion including Moses Harris, Philipp Otto Runge, George Field, Charles Hayter, J.L.F. MÉRIMÉE, Michel-Eugène Chevreul, Albert Munsell, Johannes Itten, and Josef Albers [7–21]. In this context, the ontological focus of RYB color remained substance (pigment color) and, in particular, specific types of pigment color such as gouache. Media such as this had the capacity to create a large gamut of color nuances via intermixture of the RYB triad as well as flat areas of opaque color with “hard” edges suitable for professional design output. The RYB color model was widely applied across multiple industry sectors up until the emergence of digital color technologies in the twentieth century.

The RYB color model did not feature a methodology to identify and communicate color information across different applications or industry sectors. Initially, color matching was conducted manually, but the unreliability of this process proved problematic for those applying color across diverse applications due to color mismatch issues. As a result, color notation systems were developed to aid identification and specification of

color including the Munsell color notation and color atlas (initially published 1905, 1913), the RAL color system (launched 1927), the Pantone PMS system (launched 1963), and the NCS color system (launched 1979) [22–24].

RYB Color Model

The RYB color model often reflects the format of other color models such as RGB and CMY. However, each of these color models is distinctly different, in terms of ontology and epistemological traditions, and should not be conflated or used substitutionally in any way. RYB color emerged from traditional color theory and relates to substance (pigment) color while RGB color relates to spectral profile color and visual perception of color; meanwhile, the CMYK model emerged from the way in which color is printed using various techniques common in the printing industry.

The RYB model featured in Fig. 1 is illustrated in Venn diagram format to differentiate it from other color models, whereby three exemplar pigment colors (red, yellow, and blue) represent primary and distinct components (x , y , and z), isolated via gray boundaries. Three secondary exemplar pigment hues (orange, green, and

purple) are achieved through intermixture of primary pigment hues ($x + y$, $y + z$, and $z + x$). These differentiated component sections can be considered *minterms*, or a logical conjunction of the three distinct primary component sections. As such, these minterm component sections do not represent a semitransparent overlay of primary hues arising from additive mixture of hues as per the RGB color model. Instead, the minterm components represent representative hues which are intended to represent a multitude of the potential intermixed hue nuances of orange, green, and purple. The central component represents a tertiary color achieved via intermixture of all three primary pigment hues ($x + y + z$). This tertiary color appears a murky grayish brown but can vary considerably in hue and value depending on the proportional intermixture of primary pigment hue colors.

As with similar color models, the RYB color model is a highly simplistic conceptual model which is not intended to illustrate the complexity of color or the multitude of hue nuances achievable via pigment color intermixture but a simple illustrative model of basic color creation and intermixture.

Cross-References

- [Chevreul, Michel-Eugène](#)
- [Color Combination](#)
- [Color Harmony](#)
- [Itten, Johannes](#)
- [Mayer, Tobias](#)
- [Munsell, Albert Henry](#)
- [Primary Colors](#)
- [Runge, Philipp Otto](#)

References

1. Green-Armytage, P.: The value of knowledge for colour design. *Color. Res. Appl.* **31**, 253–269 (2006)
2. O'Connor, Z.: Traditional colour theory: a review. *Color. Res. Appl.* (in press)
3. Gage, J.: *Colour and Meaning: Art, Science and Symbolism*. Thames & Hudson, London (2000)

4. Kuehni, R.G., Schwarz, A.: *Color Ordered: A survey of color order systems from antiquity to the present*. Oxford University Press, Oxford (2008)
5. Mayer, T.: Treatise on the relationship of colors. In: Forbes, E.G. (trans.) *Tobias Mayer's Opera Inedita: The First Translation of the Lichtenberg Edition of 1775*, New York (1971). http://www.gutenberg-e.org/lowengard/C_Chap55.html. Accessed 15 May 2020
6. Mayer, T.: On the relationships between colors (1775), translated by A. Fiorentini. *Color. Res. Appl.* **25**, 66–74 (2000)
7. Harris, M.: *The Natural System of Colours* (1766), Facsimile Edition. Whitney Library of Design, New York (1963)
8. Field, G.: *Chromatics: An Essay on the Analogy and Harmony of Colours*. A.J. Valpy, London (1817). https://archive.org/details/gri_33125012258899. Accessed 2 June 2020
9. Field, G.: *Chromatography: a Treatise on Colours and Pigments*. Charles Tilt, London (1835). https://archive.org/details/gri_c00033125008687523. Accessed 2 June 2020
10. Hayter, C.: *A New Practical Treatise on the Three Primitive Colours*. John Booth, London (1826). <https://archive.org/details/newpracticalrea00hayt>. Accessed 10 June 2020
11. Mérimée, J.F.L.: *De La Peinture a L'Huile*. Mme Huzard, Paris (1830). https://archive.org/details/bub_gb_NYSIPn0n38cC/. Accessed 2 June 2020
12. Mérimée, J.F.L.: *The Art of Painting in Oil and in Fresco*, English Translation by W.B.S. Taylor. Whittaker and Co, London (1839). <https://archive.org/details/artpaintinginoi01taylgoog/>. Accessed 4 June 2020
13. Chevreul, M.E.: *The Principles of Harmony and the Contrast of Colours*, English Translation by C. Martel. Kessinger, Whitefish (1855)
14. Munsell, A.H.: *A Color Notation*. Ellis, Boston (1905). <https://archive.org/details/acolornotation00munsgoog>. Accessed 24 Apr 2020
15. Munsell, A.H.: *Color Balance Illustrated: an Introduction to the Munsell System*. Ellis, Boston (1913). <https://archive.org/details/colorbalanceillustrated00munsiala/>. Accessed 5 Aug 2020
16. Munsell, A.H.: *Atlas of the Munsell Color System*. Wadsworth, Howland & Co, Malden (1915). <https://archive.org/details/AtlasMunsellcol00Muns>. Accessed 24 Apr 2020
17. Itten, J.: *The Art of Color* (1961). Wiley, New York (revised edition, 1973)
18. Itten, J.: *Design and Form: the Basic Course at the Bauhaus and Later*. Wiley, New York (1975)
19. Albers, J.: *The interaction of color*. Yale University Press, New Haven (1963)
20. Ruskin, J.: *The laws of Fésole: A Familiar Treatise on the Elementary Principles and Practice of Drawing and Painting, As Determined by the Tuscan Masters; Arranged for the Use of Schools*, vol. 1. George

- Allen, Kent (1879). <https://archive.org/details/lawsfoleafamil03ruskgoog/>. Accessed 12 Aug 2020
21. Parkhurst, C., Feller, R.L.: Who invented the color wheel? *Color. Res. Appl.* **7**, 217–230 (1982)
 22. Munsell Color Company: Atlas of the Munsell Color System. Munsell Color Company, Boston (1913). <https://munsell.com/wp-content/uploads/2014/03/atlas-munsell-color.pdf>. Accessed 6 May 2020
 23. Pantone: About Pantone; 2020. <https://www.pantone.com/about/about-pantone>. Accessed 6 May 2020
 24. NCS: NCS – our history; 2020. <https://ncscolour.com/about/our-history/>. Accessed 6 May 2020

RYB Color Theory

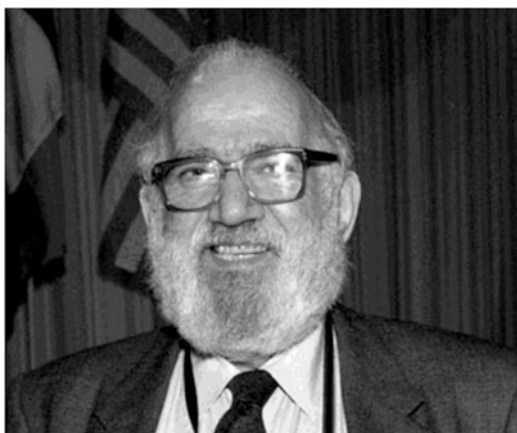
► [Traditional Color Theory](#)

RYB Triad

► [RYB Color](#)

Saltzman, Max

Ellen C. Carter
Color Research and Application and, Konica
Minolta, Pennsville, NJ, USA



Max Saltzman was a chemist, educator, and scholar. As a steadfast member of the color community in the USA, he often mentored colleagues with thoughtful and generous encouragement, which resulted in him having a profound influence in the field of color science in both industry and the art community.

Saltzman received a B. S. degree in chemistry from the College of the City of New York in 1936. In this early period he worked in the field of medical instrument research and during the war

years was in civilian service with the Chemical Warfare Service. Following the war he settled into his first career in color by joining Harmon Colors, which later became Allied Chemical Corporation, where he worked in various research and management capacities for 26 years. As a native New Yorker, he lived in the city and reverse commuted to New Jersey for this entire time period.

At Harmon he became an expert in pigments, dyes and resins by identifying the colorants, making objective color measurements, defining criteria for selection of colorants for specific applications and use of computers for color matching. We can see his interest and influences through the organizations which he joined: The American Chemical Society, the American Association of Textile Chemists and Colorists, the Inter-Society Color Council, the Society of Plastic Engineers, the Optical Society of America, the Society of Dyers and Colorists (U.K.), the Color Group (Great Britain), the Dry Color Manufacturers' Association, and the Federation of Societies for Coatings Technology.

During this time, Saltzman realized there was a need for academic and industrial training in color technology. He encouraged Dr. Walter Bauer, Dean of the College of Science at Rensselaer Polytechnic Institute to establish a new research laboratory at RPI and to recruit Dr. Fred W. Billmeyer, Jr., to run the undergraduate and graduate program in color science and to expand it to include summer short courses for people from industry. The very popular textbook, *Principles of Color Technology* by Fred W. Billmeyer and Max

Saltzman went through two editions during this period (1966 and 1981) and followed Saltzman's teaching principles of going back to the basics for a thorough understanding of the issue. Saltzman served as an Adjunct Professor for most of the 20 years the Rensselaer Color Laboratory was in existence.

Upon retiring from Allied Chemical in 1973, Saltzman left New York, moving to the west coast where he found time to follow one of his other major interests: the study and identification of ancient dyestuffs in textiles. He began by establishing a color identification laboratory in the Institute of Geophysics and Planetary Physics at the University of California. His research led him to far and wide, for example, he traveled to Peru to collect authentic samples of dyes used in ancient Peruvian textiles. This soon led to his second career in the area of conservation of historical textiles and other objects. He became active in the American Institute for Conservation of Historic and Artistic Works and the International Institute for Conservation of Historic and Artistic Works. In 1984, he was invited to present the George L. Stout Memorial Lecture, a prestigious honor from the American Institute for Conservation of Historic and Artistic Works. When the J. Paul Getty Conservation Institute was established, Saltzman became a consultant there.

Although often working behind the scenes with individuals, Max Saltzman became an internationally recognized authority in the field of color science, receiving the Armin J. Bruning Award (1969) and the Inter-Society Color Council Macbeth Award (1986). Among his many recognitions and awards, the Inter-Society Color Council also honored Max Saltzman's lifetime of contributions by bestowing on him the 2001 Godlove Award. Sadly, he died before the ceremony to receive the award.

Further Reading

1. Billmeyer, F.W., Jr.: Principles of Color Technology, 2 edn. John Wiley & Sons, New York (1966, 1981)
2. English translation of Practical color measurement: a primer for the beginner, a reminder for the expert by Anni Berger Schunn (1994)

3. Saltzman, M.: Color matching via pigment identification. *Dyestuffs*. **43**(3), 57–65 (1959)
4. Saltzman, M., Keay, A.M., Christensen, J.: The identification of colorants in ancient textiles. *Dyestuffs*. **44**, 241–251 (1963)
5. Saltzman, M., Keay, A.M.: Colorant identification. *J. Paint. Technol.* **39**(509), 360–367 (1967)
6. Saltzman, M.: The Identification of Dyes in Archaeological and Ethnographic Textiles in *Archaeological Chemistry – II*, pp. 172–185. ACS Publications, Washington, DC (1978)
7. Billmeyer Jr., F.W., Kumar, R.: Saltzman m.: Identification of organic colorants in art objects by solution spectrophotometry: pigments. *J. Chem. Educ.* **58**(4), 307 (1981)
8. Billmeyer, F.W., Saltzman, M., Kumar, R.: Identification of organic pigments by solution spectrophotometry. *Color. Res. Appl.* **7**, 327–337 (1982)
9. Saltzman, M.: Analysis of dyes in museum textiles or You can't tell a dye by its color. In: McLean, C.C., Connell, P. (eds.) *Textile Conservation Symposium*, pp. 39–1986. Los Angeles County Museum of Art, Honor of Pat Reeves Los Angeles (27)
10. Saltzman, M.: Identifying dyes in textiles. *Am. Sci.* **80**, 474–481 (1992)

Sampling Problems in Computer Graphics

Kartic Subr

Department of Computer Science, University College London, London, UK

Definition

The term “sampling” is ambiguous since it lends itself to a multitude of interpretations across various fields of study. In computer graphics, sampling problems arise in two distinct contexts. First, continuous functions (signals) are discretized via sampling. For example, a pixelated image represents an underlying continuous image signal sampled at the grid of pixel locations. The general problem is then of reconstructing a high-fidelity approximation of the continuous signal using the discretization. This application of sampling is heavily inspired by the wealth of literature in the digital signal processing community [1]. The second category of sampling problems in

computer graphics applications uses sampled values to estimate general characteristics of the sampled population [2]. For example, in Monte Carlo image synthesis, samples of the radiance incident at each pixel are used to estimate the expected radiance at the pixel, using Monte Carlo integration. This article addresses the relevance and application of each of these classes of sampling problems across various topics in computer graphics.

Two Categories of Sampling Problems

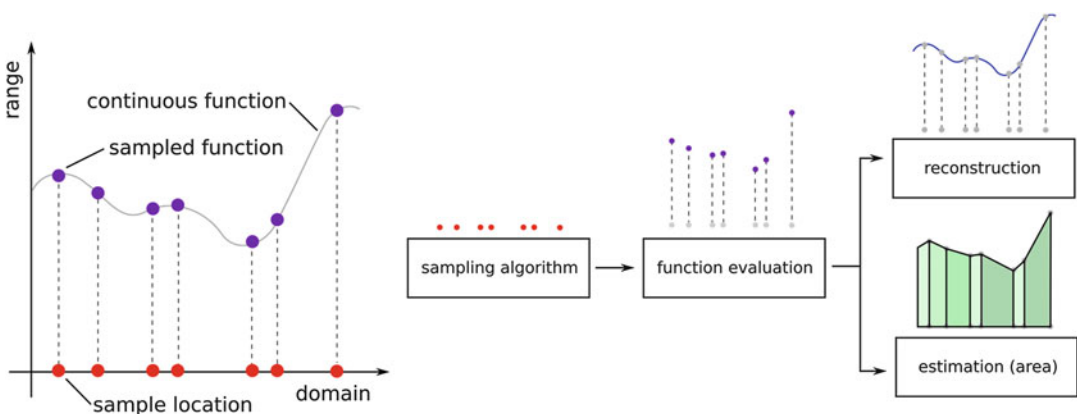
The core problem of sampling a continuous domain involves the selection of a finite set of points in the domain. The sample locations may be chosen by a deterministic algorithm or a stochastic algorithm. The evaluation of a function, defined over this domain, at these sample locations is commonly known as sampling the function. Numerous strategies exist for sampling. The relative benefits of these strategies hinge on the use of the resulting samples.

In computer graphics, the functions are sampled for two purposes (Fig. 1). First, sampled functions are sometimes the only tractable representation for the function under consideration. For example, images are rendered by computing the

amount of light at each pixel on a grid. The second purpose of sampling is to glean key attributes of the function from its sampled values. For example, the expected amount of incident light (radiance) at each pixel is estimated by sampling the infinite space of light paths arriving at that pixel from virtual light sources in the scene. Another example of statistical sampling in graphics is in usability testing, for subjective evaluation, such as rating intuitive interfaces, judging the relative quality of images, etc. This article presents an overview of sampling problems that arise in computer graphics. In addition, it briefly describes the approaches adopted for analysis and assessment of the sampling algorithms. Regardless of the use of the sampled function, the representation of the function by evaluations at a finite set of locations introduces errors. A systematic characterization of these errors is important for assessing and comparing sampling strategies.

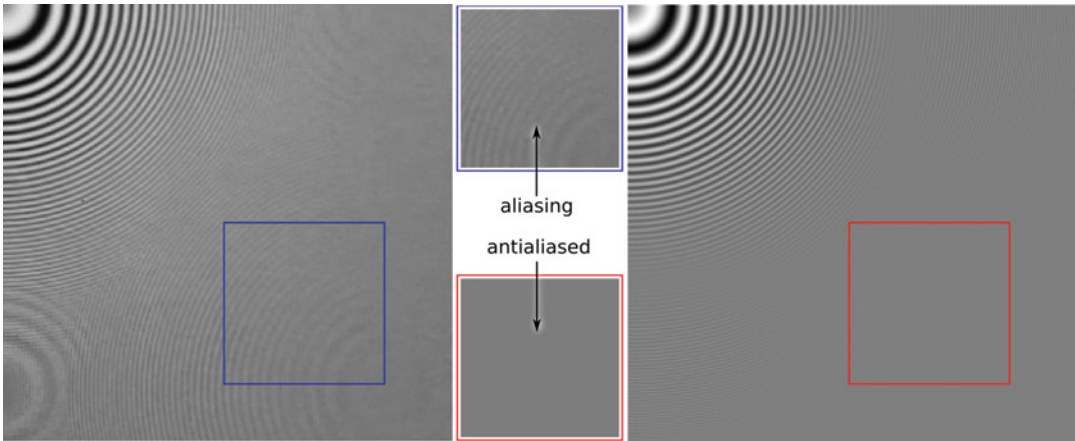
Discretization: Sampling for Reconstruction

Discrete representations are ubiquitous in computer graphics: Images are represented by a spatial grid of locations; the distribution of radiant light energy across wavelengths is often represented



Sampling Problems in Computer Graphics, Fig. 1 Sampling is the process of identifying a set of points (red circles) in the domain of a function. Sampled functions are function values (purple circles) evaluated at the sampled points. In computer graphics, sampled

functions are used for two general purposes: reconstruction of the function (blue curve) and for estimation of characteristics of the function, such as the area under its curve (e.g., the green shaded area)



Sampling Problems in Computer Graphics, Fig. 2 The errors, or artifacts, due to the discretization of a continuous function is called *aliasing*. The zone plate function (over a quadrant) is depicted here, sampled on a

256×256 regular grid. The finite set of samples introduces apparent patterns (*left*) which can sometimes be removed, or reduced, by specially designed reconstruction algorithms (*right*). Zoomed insets are shown in the *middle*

using three pixel colors (red, green, and blue); 3D geometry is commonly expressed as simplicial complexes (e.g., triangle meshes); fluid and other dynamical simulations are performed by propagating physical attributes at discrete locations. A fundamental advantage of discrete representation is that a finite set of data points, or samples, may be used to represent the function under consideration. The process of selection of the locations of these data points in the function domain is called sampling.

The distortion exhibited by the discrete function, due to an insufficient set of data points, is called aliasing. The practical elimination of this error, in general, is still considered an open problem. A thorough discussion of the fundamental problems related to sampling, reconstruction, and aliasing may be found in textbooks on signal processing [1]. The problem of anti-aliasing, or abating the distortion, has enjoyed much attention across several domains of computer graphics.

Images

The choice of a discrete grid, of pixels, to represent an image comes with a penalty – aliasing. If the underlying continuous image exhibits rapid spatial (or temporal) variation, unless the resolution of the spatial grid (resp. frame rate) is large enough, artifacts appear in the image as shown in

Fig. 2. To avoid aliasing, the sampling frequency must be twice the maximum Fourier frequency in the continuous image.

In the frequency domain, if there exists a finite frequency above which the Fourier amplitude spectrum of the image (before sampling) is zero, then aliasing can be avoided by suitably raising the number of samples on the grid. Such signals (images) are known as band limited, and this approach to anti-aliasing is called super-sampling. If the sampling rate is above twice the maximum frequency, then an ideal low-pass reconstruction filter [3] may be designed that is guaranteed to eliminate aliasing. The primary advantage of super-sampling is its simplicity and flexibility. It is widely used in rendering since it can be used in conjunction with complex primitives such as polygons, implicit functions, or parametric surfaces. Common strategies used to generate the sample locations include regular grid sampling, stratified random (jittered) sampling, and sample sets with blue-noise spectral characteristics.

If the underlying continuous signal contains positive energy at arbitrarily high frequencies or is not band limited, then no finite resolution of the discretization can eliminate aliasing. This is typical, since the existence of any step edge implies arbitrarily high frequencies. In such cases, anti-aliasing may be performed by first blurring, or

band limiting, by convolving with a smooth prefilter. Prefiltering is a way of enforcing the image signal before sampling to be band limited. Efficient prefiltered anti-aliasing is tricky for rendering: It involves continuous area integrals with detailed boundary information to find the integration limits; its applicability depends on the primitive. However, it is effectively used in texture mapping, when a high-resolution texture is available, and needs to be mapped on an object that potentially projects onto a small area in the image. Rather than mapping the high-resolution texture and then anti-aliasing at runtime, a popular approach is to query the appropriate level, based on the projected area, from a precomputed, prefiltered hierarchy of the texture. Since anti-aliasing is precomputed with this mip-mapping (multum-in-parvo-mapping) approach, it is widely used in real-time computer graphics applications such as rendering in video games.

Geometry

Polygon meshes are a popular choice for representing 3D geometry in computer graphics. They are simple and intuitive to manipulate locally, and they lend themselves to efficient rendering. Modeling or creating high-quality polygon meshes requires skilled artistry and sophisticated 3D modeling tools. To circumvent the long modeling process, artists often use 3D scans of objects as a starting point to their design. However, scanners today output point clouds as representations of underlying surfaces. The simplification and resembling of these point clouds are considered essential steps before the scanned model is usable for rendering or surface reconstruction.

Similar to the reconstruction of anti-aliased images, reconstruction of the underlying geometry from its point-sampled version is a challenging problem. The difficulty is exacerbated by the need to generalize elegant sampling theories that are applicable in Euclidean domains, such as for image processing, to arbitrary manifolds. Spectral analysis of manifolds is based on the eigenfunctions of the Laplace–Beltrami operator, just as Fourier analysis is based on trigonometric harmonics. The extension of signal processing on arbitrary

manifolds has inspired many useful algorithms for resampling point-sampled geometry, smoothing manifolds, reconstructing simplicial manifolds from point samples, etc. A comprehensive description of sampling and processing of polygonal meshes can be found in [2].

Plenoptic Functions

While rendering images with effects such as defocus (due to a finite camera aperture) and motion blur (due to a finite exposure), the radiance at each pixel is obtained by integrating estimates of incident radiance over the aperture and time. That is, at each pixel (x, y) on the image plane, radiance estimates along a direction (u, v) , so that the ray from the pixel passes through the aperture, and at times (t) are integrated.

The general rendering process involves reconstructing portions of the plenoptic function by estimating radiance at 5D sampled locations (x, y, u, v, t) . The radiance estimate at each sampled location is obtained via numerical integration in a high-dimensional space, accounting for different transport phenomena such as reflection, refraction, scattering in participating media, and occlusion. Hence, the cost of sampling the plenoptic function is high. Although the domain spans a higher-dimensional space, standard signal processing results are still directly applicable. The search for efficient reconstruction algorithms for the portions of the plenoptic function required for rendering has spawned a recent flurry of research. A recent trend in high-quality rendering is to develop ways of approximating frequency domain characteristics of the plenoptic function, using which adaptive sampling and reconstruction filters may be derived.

Physically Based Simulation

The goal of physically based simulation, in computer graphics, is to animate virtual objects by mimicking their physical interactions with the environment. The simulation may be of rigid bodies, deformable objects, or fluids. Although the discretization plays a critical role in deciding the quality – tractability, stability, or fidelity (realism) – of the simulation in all three cases, it is particularly important in fluid simulations.

The introduction of a non-penetration constraint in rigid body simulation introduces discontinuities in attributes such as position and velocities at the moment of collision. Thus, accurately identifying the moment of collision is important. Many methods have been proposed to perform collision detection, and they resort to different discretization [4] of the geometry and time domain.

Deformable objects are simulated using a number of techniques: mass-spring systems, the finite difference method, the finite element method, mesh-free methods, coupled particle systems, and reduced deformable models based on modal analysis. Each of these methods relies on a discretization of the geometry and time, and the choice of discretization has a significant effect on the stability, sensitivity, and fidelity (realism) of the simulation. The list of sampling schemes used in this context includes a cubic octree hierarchy, non-nested hierarchy of meshes (FEM), regular spatial grids (FDM), uniformly distributed samples, and samples with blue-noise spectral profiles. The interested reader is referred to [5, 6] for a survey of the methods.

The central problem in fluid simulation for computer graphics applications is the robust solution of the incompressible Navier–Stokes equations [7]. One of three strategies is popularly adopted to decide the locations where physical attributes such as pressure, viscosity, density, and velocity are iteratively computed at each time step. The first option is a regular grid over the spatial extent of the fluid and is an Eulerian approach to the problem. The second option is to represent the fluid using discrete particles that measure these attributes as the particles themselves are advected about. The third option is to use a combination of irregularly spaced samples and regular grid. As with reconstructing images and geometry, the sampling rate governs the largest frequency variation that can be faithfully reconstructed. Simply increasing the number of samples in a fluid simulation quickly results in prohibitively expensive computation times, and so adaptive sampling strategies are key to

practicable fluid simulation. Octree-based sampling, blue-noise sampling, and regular (grid) sampling are commonly used.

Statistical Sampling: Sampling for Estimation

Another application of sampling, in computer graphics, is to estimate the statistics of a function under consideration from values of the function at sampled locations. An estimator is a combination of the sampling strategy and the process of combining the samples. The fidelity of the mean and the variance of estimates produced by an estimator are both typical measures of quality for estimators. For example, a Monte Carlo estimator for the expected value of a function is the average of the function evaluated (sampled) at uniformly random locations in the domain. This simple estimator converges to the correct answer, in the limit of sampling the domain.

Physically based image synthesis involves estimation of a portion of the plenoptic function, or incident radiance at each (x, y, u, v, t) , where (x, y) is on the image plane. The problem is addressed by sampling at two levels: designing a high-level rendering strategy or estimator and then choosing low-level, specific sampling strategies.

The rendering strategy governs how light paths are sampled. For example, in path tracing, each estimate of the radiance is obtained by sampling the contribution of a single path from the space of light paths originating at (x, y) on the image plane and terminating at a light. In distributed ray tracing, incident radiance is recursively estimated at each bounce, using multiple incident samples. Metropolis light transport adopts a Markov-chain model to sample in path space, given a pool of initial paths using which the search is seeded. The specific sampling strategy defines how samples are generated. For example, at each bounce (whether path tracing or distributed ray tracing), the outgoing directions may be generated using quasi-Monte Carlo, random, stratified, or importance sampling schemes [8].

Error and Assessment

The graphics literature is rich with numerous sampling algorithms. Ideally, the invention of a core sampling algorithm develops in four stages: motivation for a novel algorithm, the design of the sampling procedure or algorithm, analysis of the proposed algorithm, and finally assessment of the resulting samples. This section describes typical approaches for the latter two stages – and assessment. The motivation and design of specific sampling algorithms are closely tied to the specific application and beyond the scope of this article.

Sampling for Reconstruction

The analysis of sampling algorithms (for reconstruction) is typically performed in the frequency domain. Fourier analysis is typical for Euclidean domains such as images and 4D plenoptic functions. For frequency analysis over spherical domains and spaces of rotations, spherical harmonics are the popular choice in computer graphics. Using the eigenfunctions of the Laplace–Beltrami operator as a frequency basis is gaining popularity for frequency analysis on arbitrary manifolds. Regardless of the choice of frequency bases, the sampling theorem from signal processing is a fundamental test to analyze the sufficiency of sampling. However, most frequency domain analyses are global. For example, it is impossible to use the Fourier transform of an image to predict local regions with high (or low) variation. A typical trick for improving local analysis is to use windowed transforms. However, any gain in localization due to windowing is at the cost of precision in the frequency domain, in accordance with the uncertainty principle.

The assessment of reconstruction error is domain specific. Generic measures such as root mean squared error or peak signal to noise ratio, although common, are typically insufficient. For example, the “quality” of reconstructed images relies heavily on human perception. Although many studies have been performed in this area, automatic assessment of reconstructed images,

including perceptual considerations, is still an open problem. For 3D geometry, the problem of automatic assessment of final appearance is compounded by its dependence on view and shading.

Sampling for Estimation

The typical measures for the assessment of estimators are accuracy (bias) and precision (variance). Bias measures the expected discrepancy between the estimated value and the actual value. Variance measures the spread of predicted values. Bias and variance are difficult to attain simultaneously. Estimators typically strike a trade-off between the two, and their relative importance depends on the application.

Despite the quantifiable measures of bias and variance, automatic assessment of estimators is not prevalent in image synthesis. Bias leads to physically inaccurate images while variance manifests as noise across the pixels. A large bias is generally believed to be less objectionable (perceptually) than a large variance. Variance reduction schemes such as stratified sampling, importance sampling, control variates, and anti-thetic variates are commonly used in Monte Carlo rendering. The resulting estimators are typically compared by manual inspection of their resulting images given a particular virtual scene. Images are generated with increasing numbers of samples and the estimator with the least noise is considered best. In addition, a converged image using an unknown estimator is visually compared against a reference solution (with a known estimator and a large number of samples). Clearly, such comparisons are not general and only valid on the particular images that the methods are demonstrated on. The development of an automatic method for such assessment is an open problem.

Cross-References

- [Global Illumination](#)
- [Image Quality](#)

References

1. Oppenheim, A.V., Schafer, R.W.: Discrete-Time Signal Processing (Chap 4, ed. 3), (3rd edition) Prentice-Hall Signal Processing Series, Pearson (2009)
2. Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., Levy, B.: Polygon Mesh Processing (2011)
3. Thompson, S.K.: Sampling. Wiley, Hoboken (2012)
4. Witkin, A., Baraff, D.: Physically based modeling: principles and practice. ACM SIGGRAPH course. <http://www.cs.cmu.edu/~baraff/sigcourse/> (1997). Accessed 14 Feb 2013
5. Sifakis, E.D., Barbic, J.: FEM simulation of 3D deformable solids: a practitioner's guide to theory, discretization and model reduction (part 1). ACM SIGGRAPH course. <http://run.usc.edu/femdefo/sifakis-courseNotes-TheoryAndDiscretization.pdf> (2012). Accessed 14 Feb 2013
6. Nealen, A., Müller, M., Keiser, R., Boxerman, E., Carlson, M.: Physically based deformable models in computer graphics. Computer Graphics Forum. **25**(4), 809–836 (2006)
7. Bridson, R.: Fluid Simulation for Computer Graphics. A. K. Peters, Wellesly (2008)
8. Shirley, P., Marschner, S.: Fundamentals of Computer Graphics (Chap. 24). CRC Press. Boca Raton. (2009)

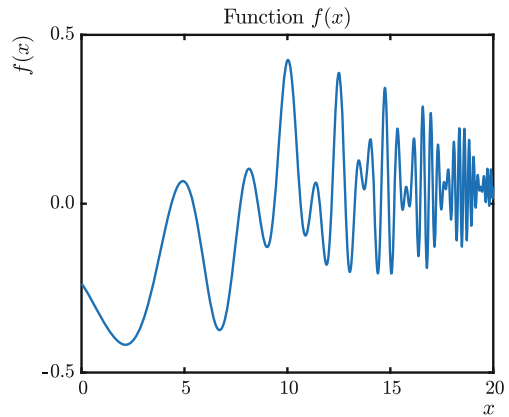
Sampling, Reconstruction, Aliasing, and Anti-aliasing

Erik Reinhard

InterDigital, Cesson-Sévigné, France

Definition

In computer graphics (as well as other disciplines), antialiasing refers to a collection of techniques to reduce the effects of aliasing. When a continuous signal or image is sampled, or a digital image is resampled, the sampling imposes a limit on the frequencies available in the input signal that can be represented in the resulting output signal. Frequencies higher than this limit (the Nyquist frequency) are mapped to lower frequencies, a process known as aliasing. In images this will be visible as objectionable artifacts. Removal of unrepresentable frequencies through low-pass filtering prior to sampling is a common anti-aliasing technique.



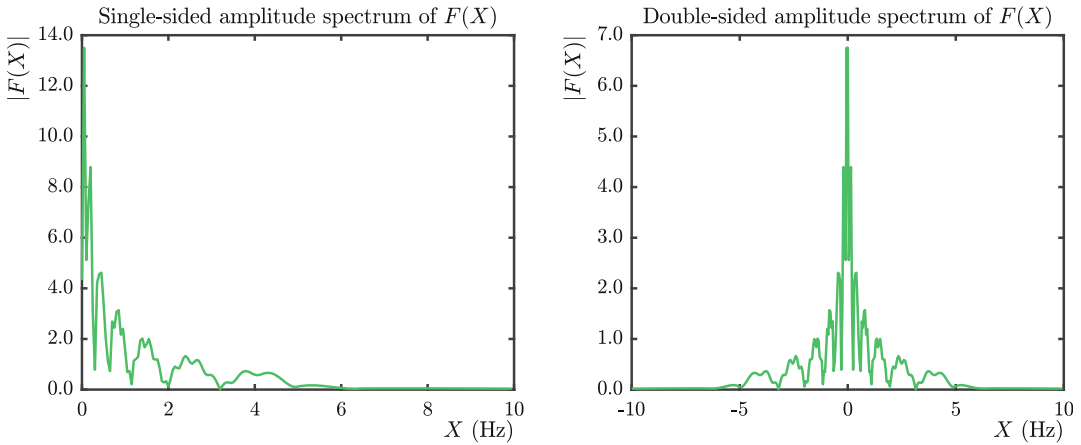
Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 1 A function containing a range of frequencies

Key Principles

A function or a signal (including images and video) typically contain oscillations at various different frequencies. A one-dimensional example is shown in Fig. 1, where from left to right the frequencies increase. For those interested in reproducing the figures shown in this article, the function displayed in Fig. 1 is given by $f(x) = (1/2) \cos(0.8\pi(x + 1)^{x/15}) \sin((\pi/40)(0.6x + 3)^{2.1})((20 - x)/20)^{0.4} + 0.05$.

Such a function can be represented digitally by sampling it. However, sampling causes a loss of information, whereby the high frequencies are lost. These are mapped (or aliased) to lower frequencies. See more on this in the “Sampling” section below.

To turn the samples back into a function, to the extent possible, a reconstruction method can be applied. This is often achieved with the aid of a filtering procedure. The filtering ensures that unrepresentable portions of the signal are removed, and thus reduce aliasing problems. Filtering can therefore be seen as an anti-aliasing method. How to achieve anti-aliasing during reconstruction is discussed in the “Reconstruction” section below. Reconstruction, for example, by filtering, is used in displays when images are displayed, but can also be performed to achieve various tasks, such as downsampling (reducing



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 2 The single-sided (left) and double-sided amplitude spectrum (right) of the function of Fig. 1

the resolution of an image) or upsampling (increasing the resolution of an image).

Filtering can also be applied to a function prior to sampling, to remove frequencies that cannot be represented by the sampled signal. This may help to reduce the effects of aliasing during reconstruction.

Sampling

A useful method to help understand the causes of aliasing is by Fourier analysis. In particular, the amplitude spectrum gives insight. The amplitude spectrum of the example function in Fig. 1 is shown in Fig. 2. The single-sided amplitude spectrum is shown on the left in this figure, and the double-sided amplitude spectrum is shown on the right. The peak of the amplitude spectrum is located at a frequency of 0 Hz. Higher frequencies are represented less. In this example, the highest frequencies are less than ten cycles per pixel (10 Hz).

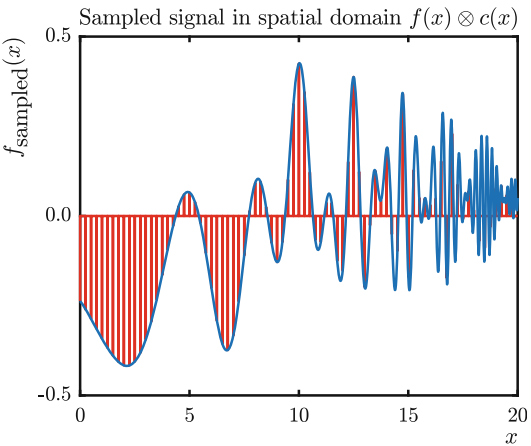
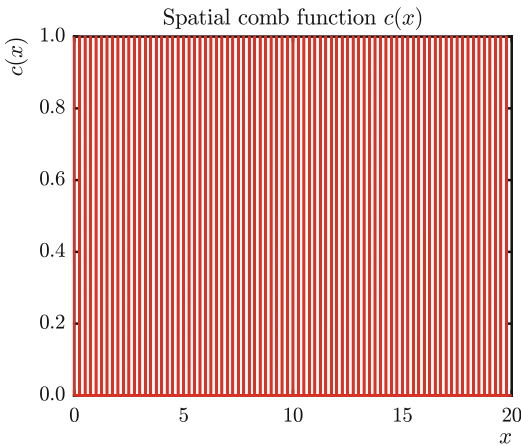
The example function of Fig. 1 is now sampled. In the spatial domain, sampling is equivalent to applying the element-wise product (or Hadamard product) of the function with a comb function. The latter is a sequence of spikes placed at the sampling locations, as shown in Fig. 3 (left). The samples are placed such that

there are four samples per unit on the x -axis. Thus, the sampling is carried out at a frequency of 4 Hz. The result of sampling is shown in Fig. 3 (right). It can be seen in this figure that the slowly varying part of the function on the left is well represented by the samples, whereas the high frequencies on the right are not well represented.

The process of sampling can be analyzed in the Fourier domain. The Fourier transform of a comb function is a comb function, as shown in Fig. 4 (left). The spikes are spaced at a rate of 4 Hz, which is consistent with the sample spacing of four samples per unit in Fig. 3.

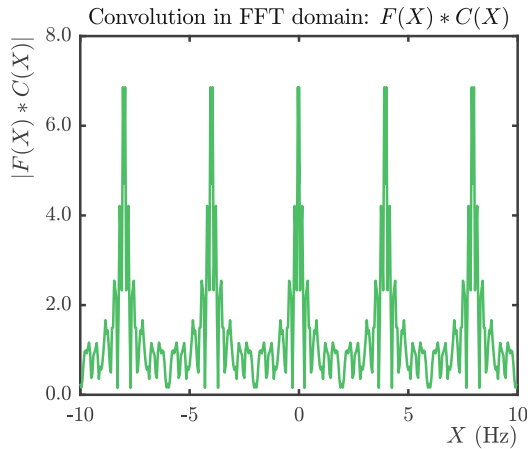
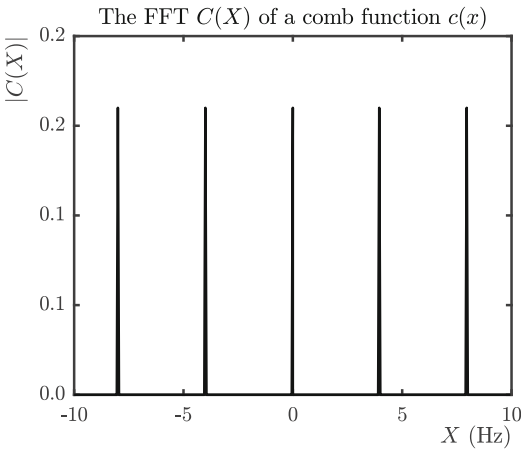
An element-wise multiplication in the spatial domain becomes a convolution in the Fourier domain. Thus, the Fourier transform of the signal (Fig. 2, right) is convolved with the Fourier domain comb function of Fig. 4 (left). This has the effect of replicating the amplitude spectrum of the function with offsets in multiples of 4 Hz, as shown in Fig. 4 (right).

The loss of information as a result of sampling comes from the fact that higher frequencies of the central copy overlap with the replicated copies. Likewise, the high-frequency tails of the replicated copies overlap with the low-frequency part of the central copy. This means that high-frequency content is misrepresented, and occurs in the sampled signal as low frequencies. The mapping of high



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 3 A comb function in the spatial domain (left). Sampling of the function of Fig. 1 is achieved by

element-wise multiplication of the function with the comb, leading to the sampled signal shown on the right



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 4 The Fourier transform of a comb function is a comb function (left). In the Fourier domain,

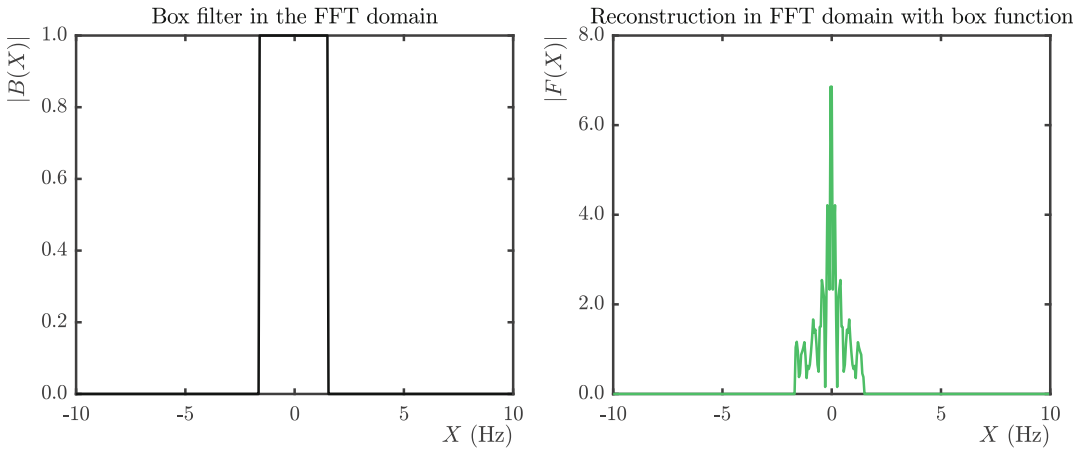
sampling is equivalent to convolving this comb with the Fourier transform of the function (see Fig. 2). The result of this convolution is shown on the right

frequencies to low frequencies as a consequence of sampling is called aliasing.

At a sampling rate of four samples per unit, the spikes in the Fourier domain are spaced at 4 Hz intervals. This means that the central copy of the sampled signal in the Fourier domain is reasonably well represented between -2 and $+2$ Hz, as can be seen in Fig. 4 (right). The highest representable frequency is therefore half of the sampling frequency. This limit is known as the Nyquist frequency.

Note, however, that the high-frequency tails of the replicated copies also overlap with the low frequencies of the central lobe, so that minor distortions of the low-frequency part of the signal can be expected.

If the function were sampled at a higher rate, then the spatial comb function of Fig. 3 would have its spikes placed more closely together. The corresponding Fourier domain comb function would have its spikes placed further apart. This would mean that the replicated copies of the signal



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 5 Ideal reconstruction in the Fourier domain involves a box function (left). Fourier domain reconstruction requires an element-wise multiplication of the sampled signal with such a box function. The box has

retained the central copy of the repeated pattern seen in Fig. 4, retaining all the frequencies that well represented in the sampled signal, and removing all the frequencies above the Nyquist frequency. This is theoretically optimal and leads to the reconstruction shown in Fig. 6

would see less overlap, and therefore higher frequencies could be represented. To ensure no overlap, the sampling frequency should be at least twice the signal frequency. In that case, the sampled signal would accurately represent the function. In the example, this would mean at least ten samples per unit, or a sampling frequency of 10 Hz.

For images, this solution could for instance involve taking more than one sample per pixel, and averaging these samples would produce a pixel value that is more representative of the function over the area that is covered by the pixel.

Increasing the sampling rate is of course a data-intensive method to limit the destructive effects of sampling. Instead, one might sample the function not at regular intervals but at random positions. In such a stochastic sampling scheme, regular aliasing patterns will be replaced by noise which is less objectionable to the human eye. A good distribution for stochastic sampling would be the Poisson disk distribution, as this guarantees a minimum distance between any pair of samples [1]. Jittered sampling first creates a regular grid, and within each grid cell a sample is placed randomly. Other random sampling strategies include N-rooks sampling and multijittered sampling [2].

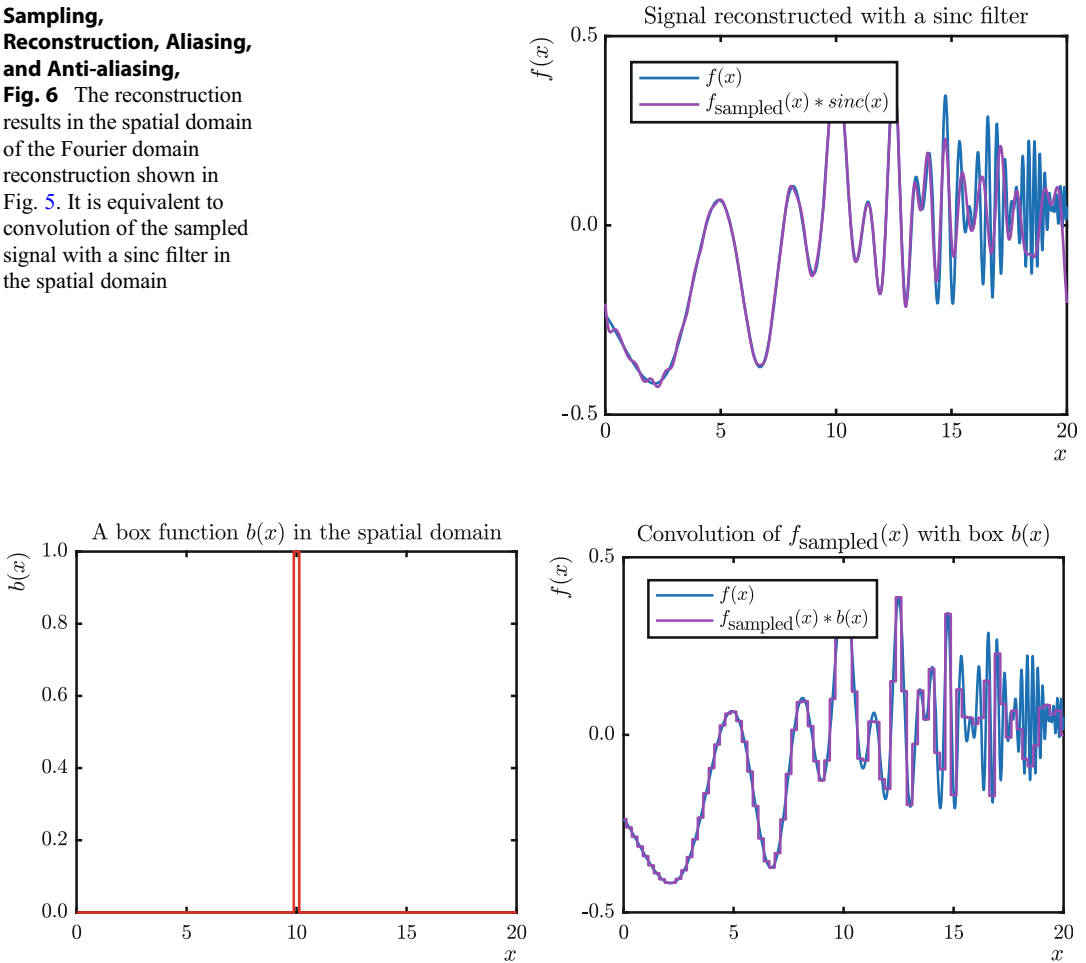
Reconstruction

In the Fourier domain representation, reconstructing the original function from the sampled signal would require the disentanglement of the central copy from the replicated copies. Given a sampling rate and associated Nyquist frequency, separating the well-represented frequencies from the ill-represented frequencies would require removing all replicated copies, leaving only the central copy. Noting that overlapping parts of the various copies cannot be disentangled, the optimal way of separating the central copy from the replicated copies in the Fourier domain is to apply a box function (Fig. 5, left), i.e., by element-wise multiplication. The result of this procedure is shown in Fig. 5 (right). Note that in this example only frequencies up to 2 Hz are retained, which is consistent with the original sampling rate of four samples per unit.

After applying an inverse Fourier transform to the result of Fig. 5 (left), the result is the reconstructed signal shown in Fig. 6. The same result could have been obtained by convolving the sampled signal with a sinc function directly in the spatial domain. Note that the low frequencies are generally reconstructed with good accuracy. The

Sampling, Reconstruction, Aliasing, and Anti-aliasing,

Fig. 6 The reconstruction results in the spatial domain of the Fourier domain reconstruction shown in Fig. 5. It is equivalent to convolution of the sampled signal with a sinc filter in the spatial domain



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 7 For computational reasons, a function is often reconstructed by convolving with a box function (left) in the spatial domain. The reconstructed result is shown on the right

high frequencies on the right, however, have not been recovered, as they have simply been truncated.

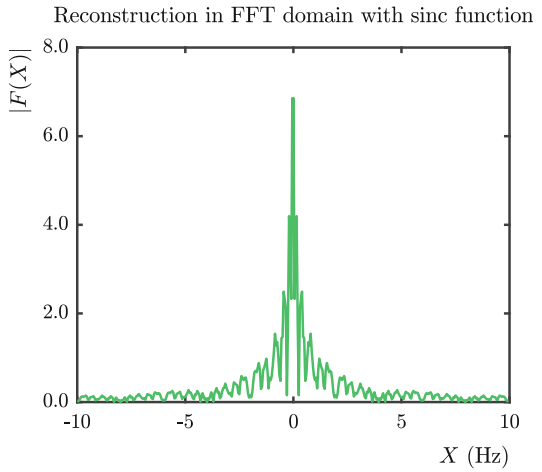
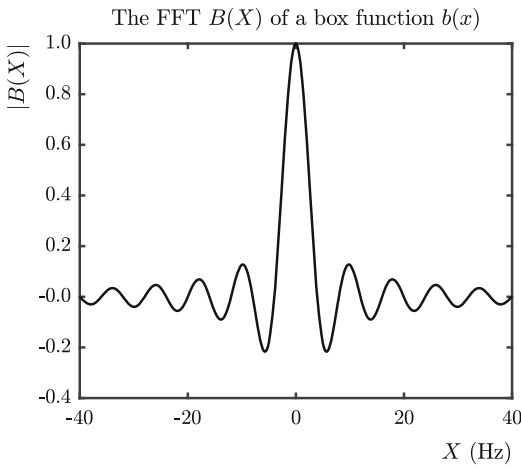
While this approach to reconstruction is optimal, as there are currently no better ways to separate representable frequencies from unrepresentable ones, the disadvantage is that this is a computationally expensive way to reconstruct a signal. While application of the box filter in the Fourier domain is not particularly expensive to compute, it requires the signal to be transformed to and from the Fourier domain, which dependent on use case may be prohibitively expensive.

Applying the convolution with a sinc filter in the spatial domain is also computationally expensive, because of the infinite extent of the sinc filter.

As a result, in practice, a variety of simpler reconstruction filters are used, with which the sampled signal is convolved in the spatial domain. Such anti-aliasing filters are discussed next.

Anti-Aliasing

Perhaps the simplest reconstruction filter is a box filter with which the sampled signal is convolved in the spatial domain. While this is the simplest of



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 8 The Fourier transform of a box function is a sinc function (left). Reconstruction with a box in the

spatial domain is equivalent to multiplication with a sinc function in the Fourier domain, as shown on the right. Compare with the two-sided amplitude spectrum of Fig. 2

the reconstruction filters, it also has the lowest quality, leading to potentially visible artifacts. Due to its simplicity, it is, however, used often.

An example of a box filter is shown in Fig. 7 (left), and the result of the reconstruction with this filter is shown in Fig. 7 (right). It can be seen here that the filter leads to noticeable steps.

The cause of this problem can be seen through Fourier analysis. The spatial box function becomes a Fourier domain sinc function, as shown in Fig. 8 (left). An element-wise multiplication of a sinc with the amplitude spectrum of Fig. 4 (right) results in the amplitude spectrum shown in Fig. 8 (right). Here, the central copy has been retained, but so have many high frequencies above 2 Hz that should have been removed.

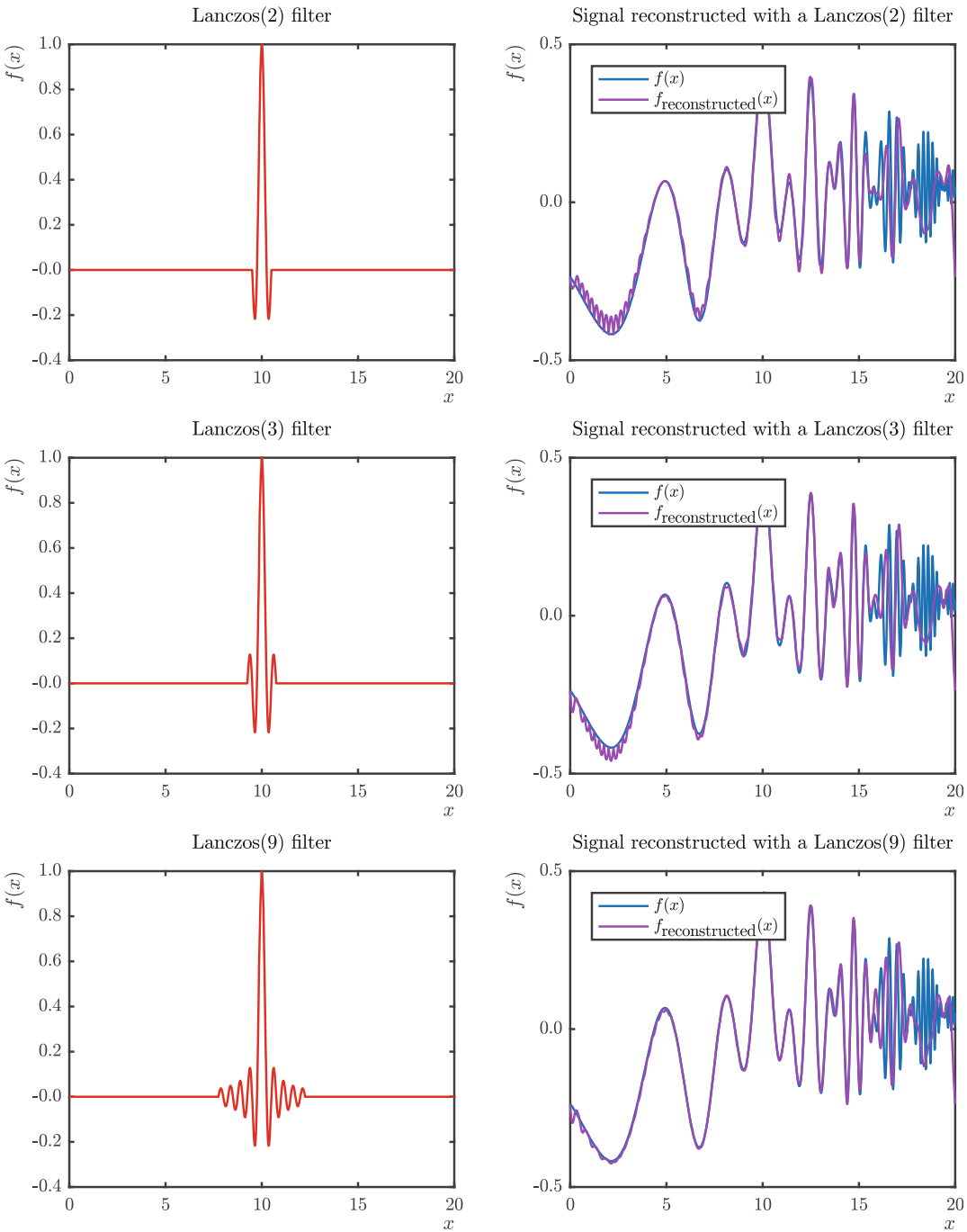
Other reconstruction filters aim to provide a better trade-off between computability and accuracy. In essence, their quality depends on how well they approximate the sinc function. Simple approximations are tent filters, cubic splines [3], cubic convolution interpolation [4], and Gaussian filters.

In particular, the Lanczos family of filters provides a reasonable balance [5, 6]. The Lanczos filter is defined as a windowed sinc function, whereby the central portion of the sinc is retained. The filter family has a single parameter which

defines how much of the central portion is retained, counted in number of (side-)lobes. Thus, a Lanczos (2) filter consists of the central lobe of a sinc as well as one side-lobe to each side. A Lanczos (3) filter includes two side lobes on each side, while a Lanczos(n) filter includes ($n - 1$) side-lobes. Examples of such filters, along with their reconstruction results, are shown in Fig. 9. As can be seen in this figure, incorporating more side-lobes provides a higher accuracy. However, the spatial convolution becomes more computationally expensive because the size of the filter kernel increases.

Image Downsampling

The theory explained above is routinely applied to images. To see the visual effect of a variety of reconstruction filters, an input image was downsampled by a factor of 8 using several reconstruction filters (nearest neighbor, bicubic, lanczos (2), lanczos (3), and box). The resulting images were then upsampled by a factor of 8 using pixel replication to arrive at the original size. The input image was chosen to be a quarter of HD size (which is 1920×1080), so that when reproduced in this document at 300 DPI, the horizontal size is 3.2 inches ($1920/2/300 = 3.2$). The images in

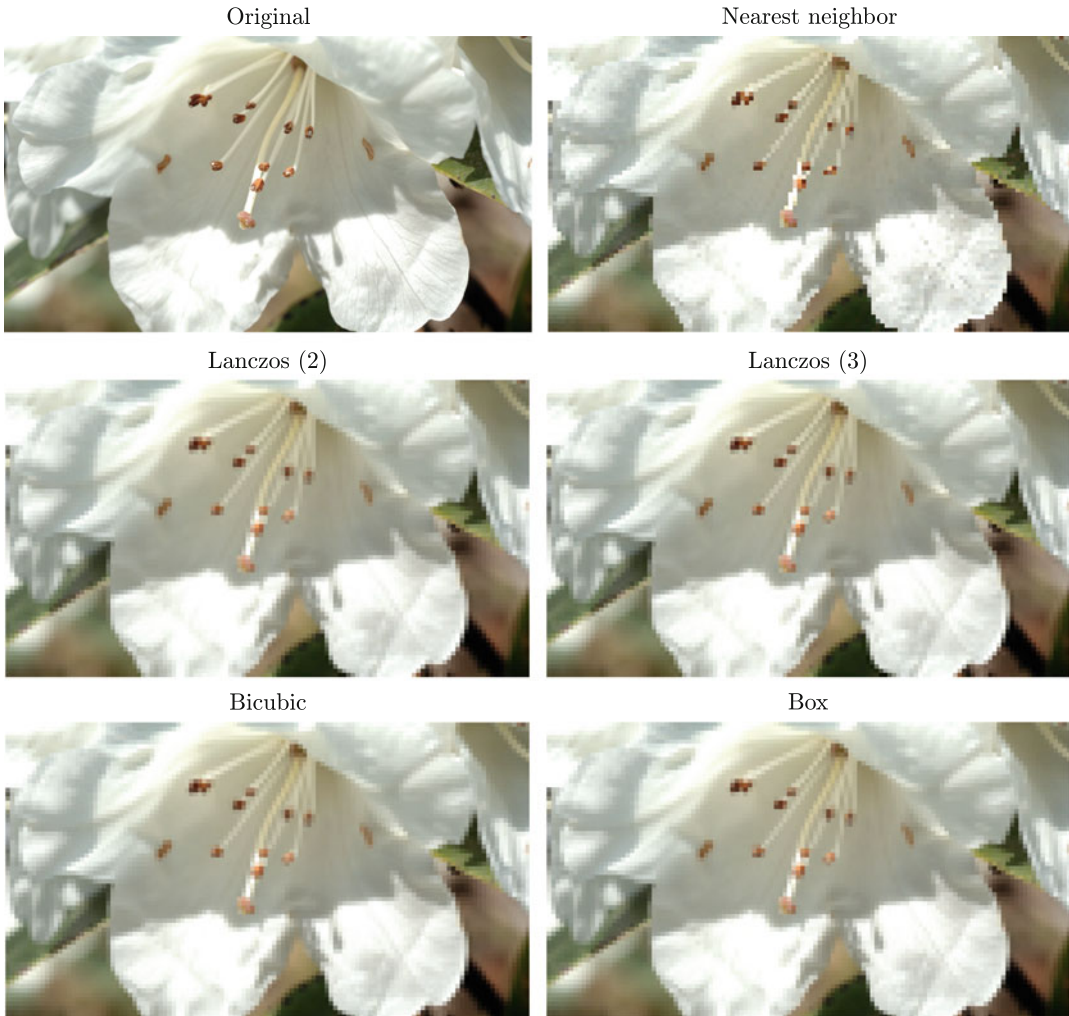


Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 9 Lanczos reconstruction filter kernels (left), and their reconstruction results (right). Top to

bottom: Lanczos kernels with parameters 2, 3, and 9 were used, leading to an increasingly accurate (but computationally expensive) signal reconstruction

Fig. 10 are therefore reproduced at 300 DPI at 3.2 inches width to avoid the typesetting software performing additional resampling.

In Fig. 10, the images were downsampled in RGB space to show the effect of various reconstruction filters. Other color spaces may be



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 10 Image downsampling by a factor of 8 using various reconstruction filters

appropriate for certain tasks. For example, color spaces that separate luminance from chrominance can be used to great advantage by exploiting the fact that human vision is less sensitive to errors in chrominance than it is to errors in luminance. Using the same downsampling and upsampling procedure outlined above, Fig. 11 shows an image processed in RGB space. Here the three color channels are treated identically (as in Fig. 10), showing a significant degradation of the image as a result of the sampling. Figure 11 also shows an image processed in YCbCr color space. The color space decomposes the image in a

luminance channel and two chromatic channels. The two chromatic channels were downsampled by a factor of 8 and then for the purpose of visualization upsampled to its original size by pixel replication.

Further Reading

This article owes much to two excellent articles written by Jim Blinn 30 years ago: “What we need around here is more aliasing,” and “Return of the Jaggy” [7, 8].

Original



YCbCr



RGB



Sampling, Reconstruction, Aliasing, and Anti-aliasing, Fig. 11 Chromatic downsampling by a factor of 8

References

1. Mitchell, D.P.: Generating antialiased images at low sampling densities. *ACM Comput. Graph.* **21**(4), 65–72 (1987). (Proceedings SIGGRAPH'87)
2. Suffern, K.: *Ray Tracing from the Ground Up*. A. K. Peters/CRC Press, New York (2007)
3. Hou, H.S., Andrews, H.C.: Cubic splines for image interpolation and digital filtering. *IEEE Trans. Acoust. Speech Signal Process.* **26**(6), 508–517 (1978)
4. Keys, R.G.: Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **29**(6), 1153–1160 (1981)
5. Lanczos, C.: *Applied Analysis*. Prentice-Hall, Englewood Cliffs (1956)
6. Duchon, C.E.: Lanczos filtering in one and two dimensions. *J. Appl. Meteorol.* **18**(8), 1016–1022 (1979)

7. Blinn, J.: Jim Blinn's corner – what we need around here is more aliasing. *IEEE Comput. Graph. Appl.* **9**(1), 75–79 (1989)
8. Blinn, J.: Jim Blinn's corner – return of the jaggy [high frequency filtering]. *IEEE Comput. Graph. Appl.* **9**(2), 82–89 (1989)

Sapir-Whorf Hypothesis

- [Effects of Color Terms on Color Perception and Cognition](#)

Schiller

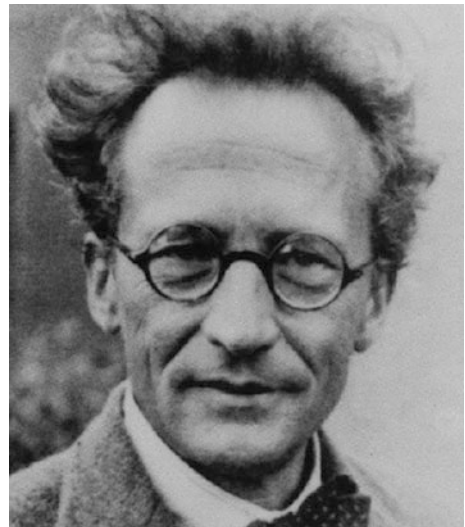
- [Iridescence \(Goniochromism\)](#)

Schrödinger, Erwin

Michael H. Brill¹ and Rolf G. Kuehni²

¹Lawrenceville, NJ, USA

²Charlotte, NC, USA



Erwin Schrödinger (1887–1961) was born in Erdberg, Austria, to a father of Austrian and a mother of mixed Austrian-English descent. He

Schrödinger, Erwin,
Fig. 1 Schrödinger's
 portrait on the Austrian
 1000-Schilling note



studied physics in Vienna under Franz Exner and became his assistant in 1911. He was influenced early by the writings of the German philosopher Arthur Schopenhauer resulting, among other things, in his interest in color theory. Three of his papers on color were written in Vienna after he concluded his military service in 1918 and before assuming a position at the University of Zürich in 1921. There he published six more papers on color and vision. His interest in color, which did not extend beyond his years in Zürich, apparently was due to his deep involvement with the works of Schopenhauer who in turn was influenced by Goethe.

In Zurich, Schrödinger did his most important work, on quantum wave mechanics, for which he shared with Paul Dirac the 1933 Nobel Prize in physics. In 1940, Schrödinger moved to Ireland, where he remained until his retirement in 1955. Then he returned to Vienna. Among his achievements is the thought experiment known as Schrödinger's Cat.

Schrödinger's contributions to color include the following:

- (a) He offered the first mathematical proof [1] of the theorem sketched by Wilhelm Ostwald that, under any illuminant, the object-color tristimulus locus is contained by a two-dimensional manifold generated by reflectances that are 1 or 0 at each wavelength and with at most two transitions between 0 and 1. Robert Luther later called these reflectances "optimal colors." The maximality of two transitions depends on the convexity of the

spectrum locus, as is clear in his very brief proof: "*If a pigment has three transition points that in the chromaticity diagram are not located on a straight line, by moving the transition points its reflectance can be changed in a manner that results in a lighter pigment of the same chromaticity coordinates. Therefore, it cannot be optimal.*"

- (b) He was the first to use differential geometry for color space, to import brightness as one of the space's coordinates, and to infer total color difference (in units of just-noticeable differences) through arclength along geodesics [2, 3].
- (c) He was the first to offer a mathematically detailed connection of the Young-Helmholtz three-color theory and Hering's opponent-color theory [4], after Helmholtz offered such a concept in 1896 and von Kries that of a zone theory in 1905.

In addition to this work, Schrödinger also published several other works relating to color science [5–9].

Besides Isaac Newton, we are unaware of any color scientist other than Schrödinger whose portrait appeared on paper money (Fig. 1). (Goethe satirized paper money [*Faust*, Part 2, Act I, Scene 4], but did not appear on it.).

References

1. Schrödinger, E.: Theorie der Pigmente von größter Leuchtkraft. *Ann. Phys.* **62**(4), 603–622 (1920)

- (translation, Theory of pigments of greatest lightness, is at <http://www.iscc.org/resources/translations.php>)
2. Schrödinger, E.: Grundlinien einer Theorie der Farbenmetrik im Tagessehen. *Ann. Phys.* **63**(4), 397–426; 427–456; 481–520 (1920) (translation, Outline of a theory of color measurement for daylight vision, is in D MacAdam, *Sources of Color Science*, MIT Press, 1970)
 3. Schrödinger, E.: Die Gesichtsempfindungen, Müller-Pouille's Lehrbuch der Physik 2/1, 11. Auflage, 456–560, Braunschweig: Vieweg (1926). (11 pages translated, Thresholds of color difference, in D. MacAdam, *Sources of Color Science*, but chapter title and length of chapter not indicated)
 4. Schrödinger, E.: Über das Verhältnis der Vierfarben- zur Dreifarben-theorie, Sitzungsberichte der Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse, Abteilung 2a. **134**, 471–490 (1925) (translation, On the relationship of four-color theory to three-color theory, is in *Color Res. Appl.* **19**, 37–47 (1994))
 5. Farbenmetrik. *Z. Phys.* **1**, 459–466 (1920) (no known translation)
 6. Schrödinger, E.: Ton und Farbe, *Neue Zürcher Zeitung*, 3. Februar, (1923) (In a Swiss newspaper)
 7. Schrödinger, E.: Über den Ursprung der Empfindlichkeitskurven des Auges. *Naturwissenschaften* **12**, 925–929 (1924) (no known translation)
 8. Schrödinger, E.: Über Farbmessung. *Phys. Z.* **26**, 349–352 (1925) (no known translation)
 9. Schrödinger, E.: Über die subjektiven Sternfarben und die Qualität der Dämmerungsempfindung. *Naturwissenschaften* **13**, 373–376 (1925) (not aware of translation)

characterized by unilateral, throbbing, and moderate-to-severe headache lasting 4–72 h, typically accompanied by nausea and photophobia and phonophobia [2]. In approximately 20% of patients, the migraine can also be preceded by transient neurological symptoms (aura) that usually develop gradually over >5 min and last for <60 min. Auras are most frequently visual (>90%), but may also involve other sensory symptoms (pins and needles, numbness), hemiparesis or speech deficits. Visual symptom (aura) often has a bimodal progression with positive symptom (**scintillating**) followed by negative symptom (**scotoma**). To date, the trigger mechanisms of migraine with aura are cortical spreading depression (CSD). Migraine with aura (MWA) is diagnosed according to the operational diagnostic criteria of The International Classification of Headache Disorders third edition [2]. Although, in migraine with aura, headache which fulfills the same features as migraine without aura (MWOA) frequently occurs following the aura, there are also cases in which, along with aging, there is no headache that fulfills the features of MWOA, or in which there is only the aura but no headache [3]. Population-based studies indicate that a diagnosis of migraine with aura is associated with an increased risk of comorbidities [patent foramen ovale [4], stroke [5], restless legs syndrome [6]].

Scintillating Scotoma

Muneto Tatsumoto

Medical Safety Management Center, Dokkyo
Medical University Hospital, Tochigi, Japan

Synonyms

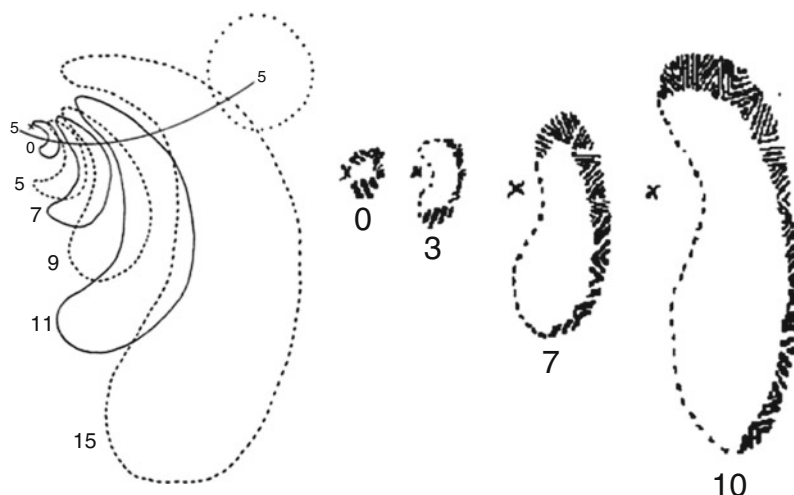
[Aura](#)

Definition

Migraine is one of the most common neurological disorders and is observed in about 11% of the world population [1]. Migraine attacks are

Overview

The patterns of the visual aura are diverse, but scintillating scotoma is representative. Scintillation is a positive symptom in which light is seen despite the lack of any input visual stimulus, and scotoma is the negative symptom in which an input visual stimulus, in the same locus as the scintillation after it has disappeared, is not visible for some time. The phenomenon of the extension of these positive (scintillation) and negative (scotoma) symptoms within the visual field is presumed to be in correspondence with changes in neural activity in the occipital lobe. It is understood that these changes in neural activity are a phenomenon known as cortical spreading



Scintillating Scotoma, Fig. 1 The analytical drawings of the spatiotemporal evolution of spreading migraine hallucinations. Left: a scintillating scotoma, spreading at the intervals shown. The blind spot is shown as the dotted circle, and the fixation point as x; alternate outlines are solid and dotted to avoid confusion; scintillations were

confined to the region above the line s-s. Right: successive forms of a scintillating scotoma, mapped at different time intervals as indicated; the fixation point is shown as x. From Lashley [7]. Copyright © 1941 American Medical Association. All rights reserved

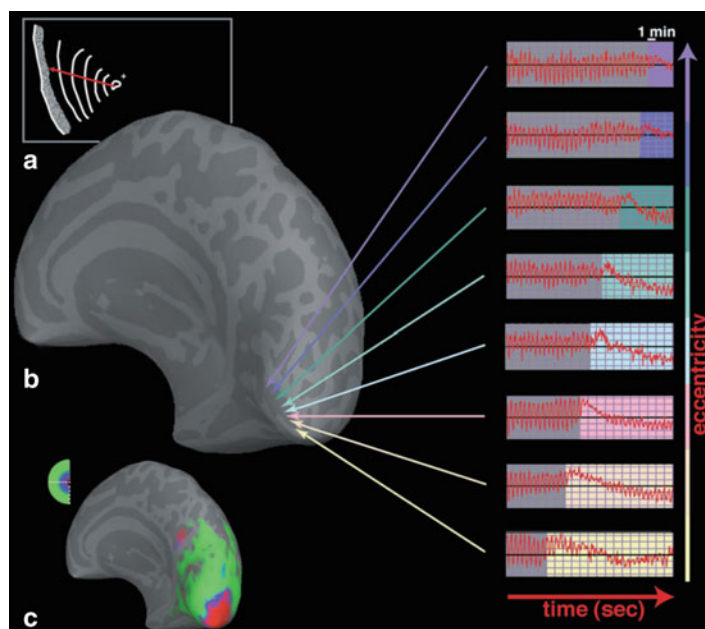
depression (CSD). CSD is a wavelike, self-propagating phenomenon of neuronal depolarization, which originates in the cerebral cortex (2–6 mm/min), as well as of inhibition of neuronal activity.

Lashley made a detailed study of his own scintillating scotoma, calculated the speed of migration of the scintillating scotoma in the visual field, and showed that it migrates at approximately 3 mm/min in the visual cortex of the occipital lobe [7] (Fig. 1).

With CSD in an experimental rabbit model, a transient inhibition of neural activity is observed upon electrical stimulation of the cerebrum, which recovers to the state of normal spontaneous activity after a certain period of time [8]. Also, when records were made by an array of recording electrodes on the cerebral surface of a rabbit, a phenomenon was observed in which transient inhibition of cerebral activity was sequentially propagated after electrical stimulation. The speed of propagation of CSD in the experimental rabbit model and the speed of propagation of the scintillating scotoma reported by Lashley were similar, both having been found to be approximately 3 mm/min. The electrophysiological

characteristics of CSD which have been indicated include (1) CSD occurs under specific conditions, (2) its extension is autonomous, (3) the speed of extension is 3 mm/min, (4) inhibition of each locus is sustained for 1–3 min, with a total refractory period longer than 1 min, and (5) all intrinsic/induced neural activities are inhibited [9].

The first evidence for the CSD hypothesis of migraine in the human cerebrum was a study conducted of regional cerebral blood flow (rCBF) in the aura phase using Xe-CT, in which spreading oligemia originating in the occipital lobe, similar to CSD, was observed [10]. The oligemia preceding hyperemia is similar to the hemodynamics, which is observed in experimental CSD. Woods et al. observed, by positron emission tomography of cases in which migraine arose accompanied by abnormal vision, spreading hypoperfusion in which, together with the onset of an episode, an area of reduced rCBF gradually extended forward from the bilateral occipital lobe [11]. The best-known report related to this is that of Hadjikhani et al. [12]. In one of the three patients, it was possible to start the test before the onset of aura since an episode was induced with near certainty upon engaging in basketball



Scintillating Scotoma, Fig. 2 Spreading suppression of cortical activity during migraine aura, as shown by functional MRI. **(a)** A drawing of the progression over 20 min of the scintillating field defect, as described by the patient. **(b)** The spread of the signal from the posterior pole towards more anterior regions representing the peripheral visual field; MR perturbations develop earlier in the foveal regions than in the periphery, consistent with the

progression of the aura from the center to the periphery. **(c)** The MR maps of retinotopic eccentricity acquired during interictal scans. As shown in the logo in the upper left, voxels that show retinotopically specific activation in the fovea are coded in red. Parafoveal eccentricities are shown in blue, and more peripheral eccentricities are shown in green. From Hadjikhani et al. [12] Copyright 2001 National Academy of Sciences, U.S.A.

practice. Upon functional MRI with presentation of visual stimulus, the signal in the visual cortex became unresponsive to stimuli as the aura began and the baseline signal slowly increased at the same time. This originated in the V3a area, extended with a speed of 3.5 mm per min to the adjacent occipital cortex including the primary and secondary visual cortices, and also coincided with the areas of the visual aura in the visual field. Subsequently, the cerebral response to visual stimuli remained weak over more than 15 min and the baseline diminished gradually to become lower than before the episode (Fig. 2).

The transient rise in this signal was similar to the change in blood flow, which was observed in accompaniment to CSD in the experimental model. This report was evidence in support of the view that the phenomenon of CSD is involved in the pathology of scintillating scotoma in migraine.

Cross-References

► [Blind Spot](#)

References

1. Robbins, M.S., Lipton, R.B.: The epidemiology of primary headache disorders. *Semin. Neurol.* **30**, 107–119 (2010)
2. Headache Classification Committee of the International Headache Society: The international classification of headache disorders, 3rd edition. *Cephalalgia*. **38**, 1–211 (2018)
3. Aiba, S., Tatsumoto, M., Saisu, A., Iwanami, H., Chiba, K., Senoo, T., Hirata, K.: Prevalence of typical migraine aura without headache in Japanese ophthalmology clinics. *Cephalalgia*. **30**, 962–967 (2010)
4. Takagi, H., Umemoto, T., for the ALICE (All-Literature Investigation of Cardiovascular Evidence) Group: A meta-analysis of case-control studies

- of the association of migraine and patent foramen ovale. *J. Cardiol.* **67**, 493–503 (2016)
5. Li, L., Schulz, U.G., Kuker, W., Rothwell, P.M.: For the Oxford vascular study. Age-specific association of migraine with cryptogenic TIA and stroke: population-based study. *Neurology.* **85**, 1444–1451 (2015)
 6. Schürks, M., Winter, A., Berger, K., Kurth, T.: Migraine and restless legs syndrome: a systematic review. *Cephalalgia.* **34**, 777–794 (2014)
 7. Lashley, K.S.: Patterns of cerebral integration indicated by the scotomas of migraine. *Arch. Neurol. Psychiatr.* **46**, 331–339 (1941)
 8. Leão, A.A.P.: Spreading depression of activity in the cerebral cortex. *J. Neurophysiol.* **7**, 359–390 (1944)
 9. Somjen, G.G.: Mechanisms of spreading depression and hypoxic spreading depression-like depolarization. *Physiol. Rev.* **81**, 1065–1096 (2001)
 10. Olesen, J., Larsen, B., Lauritzen, M.: Focal hyperemia followed by spreading oligemia and impaired activation of rCBF in classic migraine. *Ann. Neurol.* **9**, 344–352 (1981)
 11. Woods, R.P., Iacoboni, M., Mazziotta, J.C.: Bilateral spreading cerebral hypoperfusion during spontaneous migraine headache. *N. Engl. J. Med.* **22**, 1689–1692 (1994)
 12. Hadjikhani, N., Sanchez del Rio, M., Wu, O., Schwartz, D., Bakker, D., Fischl, B., et al.: Mechanisms of migraine aura revealed by functional MRI in human visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4687–4692 (2001)

Screening Plates

- ▶ [Pseudoisochromatic Plates](#)

Segmented High Beam

- ▶ [Automotive Front Lighting System \(Status 2020\)](#)

Semantic-Color Interaction

- ▶ [Color Name Applications in Computer Vision](#)

Semiconductor Nanocrystals

- ▶ [Quantum Dots](#)

Simple Colors

- ▶ [Primary Colors](#)

Simulation

- ▶ [Color Category Learning in Naming-Game Simulations](#)

Simulation and Image Synthesis of Natural Phenomena

- ▶ [Rendering Natural Phenomena](#)

Simultaneous Color Contrast

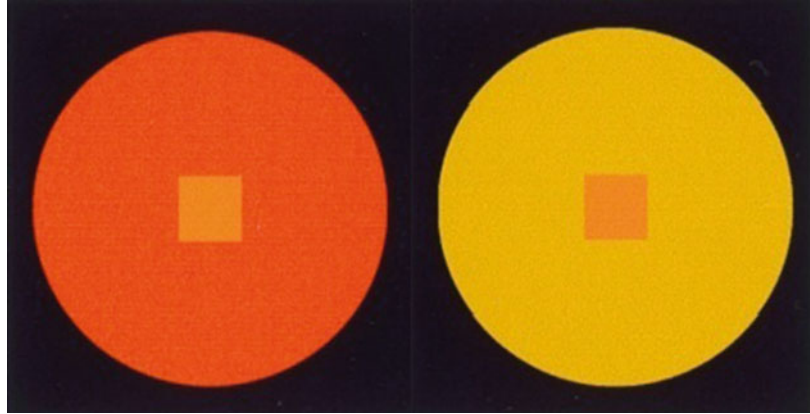
A. Soranzo
 Faculty of Development and Society,
 The Department of Psychology, Sociology and
 Politics, Sheffield Hallam University,
 Sheffield, UK

Definition

Simultaneous color contrast is the condition whereby two surfaces with the same spectral composition are perceived to have a different color when they are placed against different surrounds with different chromaticity (see Fig. 1).

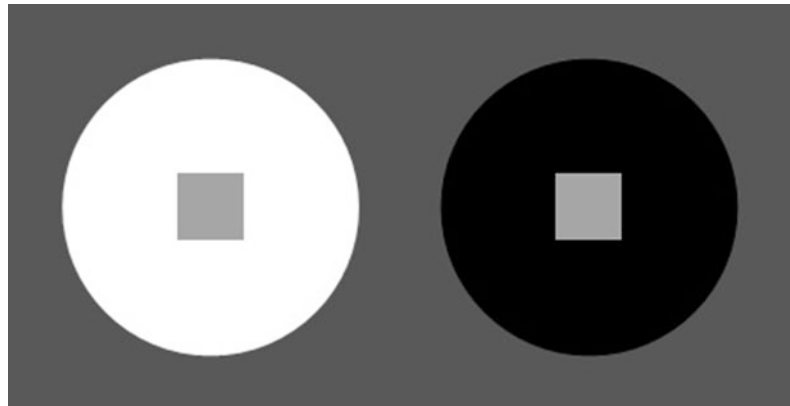
Simultaneous Color Contrast, Fig. 1

The spectral components of the squares in the centers of the red and yellow surrounds are the same. The appearance of the color is different, however. The two squares appear to acquire the opposite spectral component of their surrounds



Simultaneous Color Contrast, Fig. 2

The gray squares in the centers of the white and black surrounds are the same. The appearance is different, however: the square to the left appears darker than the square to the right



History and Theories

The term “simultaneous” was introduced by Chevreul to “distinguish this phenomenon to the ‘successive’ contrast, where two colors appear in succession upon the same retinal area” [1, p. 264].

The term “contrast” indicates that the perceived color of the surfaces is “contrasted” by the color of the surround. As this term is also used in the literature to specify the relative intensity of the stimulation, some authors prefer the term “induction” over contrast.

When the surrounded and surrounding areas are achromatic, this phenomenon is named simultaneous brightness (or lightness) contrast (see Fig. 2).

This is one of the most studied phenomena in visual perception and has been the focus of centuries of debate that has interested scientists and philosophers since Aristotle’s time [2]. In the

nineteenth century controversy raged between Hering [3], who supported an explanation based on retinal neuron interaction processes, and Helmholtz [1], who was in favor of an explanation based on higher level processes, involving assumptions about the configuration as a whole.

The retinal-based interpretation (also referred to as the “low-level” interpretation) was particularly in vogue during the 1960s mainly because of the physiological discovery of the lateral inhibition process in the limulus retina [4]. To explain the contrast phenomenon the retinal-based interpretation focuses on the notion that the receptors that are stimulated by the surround then send inhibition to the receptors that are stimulated by the surrounded area. According to this view, the simultaneous color contrast shown in Fig. 1 occurs because the receptors stimulated by the red surround inhibit the red sensitive receptors

stimulated by the orange square, leading to a yellowish appearance. Conversely, the receptors stimulated by the yellow surround inhibit the yellow sensitive receptors stimulated by the orange square, leading to a reddish appearance.

A similar explanation is provided for the contrast phenomenon seen in Fig. 2. The receptors stimulated by the light surround send inhibition to the receptors stimulated by the surrounded square causing perceptual darkening. On the other hand, the receptors stimulated by the dark surround send little inhibition to nearby receptors and, therefore, there is no darkening effect [5].

During the last few decades, however, the balance has been shifted toward more high-level theories. This is because a series of phenomena have been presented that are difficult to explain on the basis of retinal interactions. One of the best-known demonstrations is the Benary cross (see Fig. 3), discovered by Max Wertheimer and studied further by Benary [6].

Although both gray triangles are surrounded by the same amount of black and white, the left triangle appears lighter than the right triangle. Even more appealing is the contrast phenomenon found by Agostini and Galmonte [7]. In this case, a gray region surrounded by a dark area appears darker than an identical gray region surrounded by a light area (see Fig. 4).

The interested reader can find more of these effects elsewhere [8–12]. As these examples are difficult to be explained on the basis of retinal interactions, they have been interpreted according to the principles of perceptual organization [13, 14]; that is, the perceived color of a surface is determined primarily by the global contrast between the surface and the color of the surfaces to which it perceptually belongs (where perceptual belongingness refers to the grouping of a set of apparent elements into a perceived whole).

The low-level explanation has not, however, been completely abandoned as some contrast phenomena seem to involve retinal interactions such as those that generated by gradual luminance transitions (for example, [15–18]). Figure 5 shows a paradoxical contrast effect generated by



Simultaneous Color Contrast, Fig. 3 The Benary cross. The two gray triangles are the same. Although they are surrounded by the same amount of white and black area, the left triangle appears lighter than the right triangle

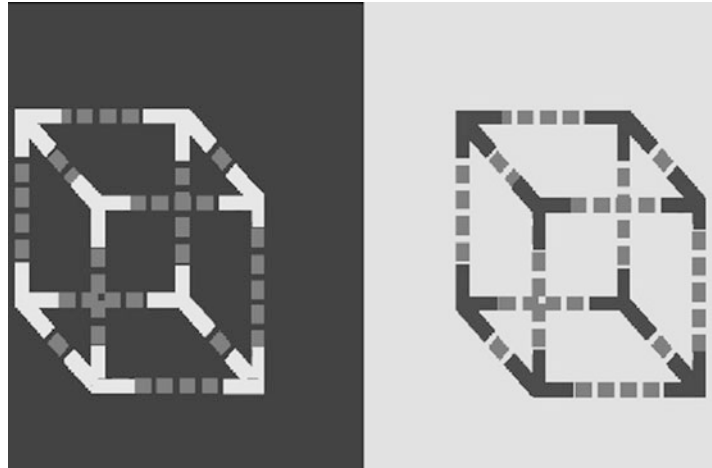
luminance transitions. The luminance surrounding the two little squares to the left side is the same, but the top square looks darker. Similarly, the luminance surrounding the two little squares to the right side of Fig. 5 is the same, but the top one looks lighter. Authors named this phenomenon the “phantom illusion” [19] because it is generated by imperceptible luminance gradients, which profile is represented next to each display.

Besides the debate between low- and high-level supporters, a vivid debate exists among scientists asserting that the retinal image is decomposed by high-level mechanisms. This is the debate between the *framework* and the *layer* types of decomposition models [27].

The framework model maintains that the visual system, in agreement with the belongingness principles, processes the light reflected by surfaces that reaches the eyes by grouping it into a set of contiguous frameworks. According to this model, contrast effects are generated by the geometric and photometric relationships among the surfaces within the same perceptual group, that is, within the same framework [14, 20].

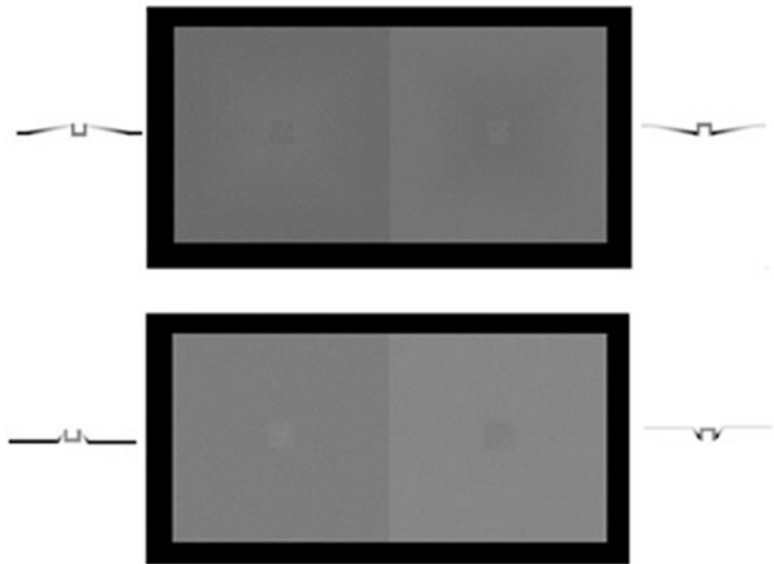
Simultaneous Color Contrast, Fig. 4

The dashed elements of the cubes are the same gray but those to the left appear darker than those to the right despite of the fact that they are surrounded by a darker area



Simultaneous Color Contrast, Fig. 5

The phantom illusion. Wide unnoticeable luminance transitions generate contrast effects on the surrounded targets (top). Narrow unnoticeable luminance transitions generate assimilation (bottom)



Conversely, the *layer* model claims that the visual system operates by splitting the light reaching the eyes into separate overlapping layers which correspond to separate physical contributions (illumination, reflectance, transparency, etc.). According to this model, contrast effects are generated when the light reaching the eyes is misattributed to the surface color and illumination. In other words, part of the light that the visual system should attribute to the surface color is instead attributed to the illumination or vice-versa [21–29].

To disentangle these two models Soranzo, Lugin, and Wilson [30] studied the contrast

phenomenon in the Virtual Reality Cave. This apparatus permitted the manipulation of the geometrical relationships between the surfaces while maintaining the same photometric relationships (i.e., the amount of light reaching the observers' eyes remained constant). However, findings indicated that both of the decomposition processes which the framework and the layer models advocate are *jointly* accountable for the color contrast phenomenon. The color contrast phenomenon may therefore be attributed to the summative effect of the framework and the layer processes.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Chevreul, Michel-Eugène](#)
- [Color Contrast](#)
- [Perceptual Grouping and Color](#)

References

1. Helmholtz, H.V.: *Helmholtz's Treatise on Physiological Optics*. Optical Society of America, New York (1866/1964)
2. Wade, N.J.: Descriptions of visual phenomena from Aristotle to Wheatstone. *Perception*. **25**, 1137–1175 (1996)
3. Hering, E.: *Outlines of a Theory of the Light Sense*. (Translated by L. M. Hurvich), D. Jameson. Harvard University Press, Cambridge, MA (1874/1964)
4. Hartline, H.K., Wagner, H.G., Ratliff, F.: Inhibition in the eye of *Limulus*. *J. Gen. Physiol.* **39**, 651–673 (1956)
5. Cornsweet, T.N.: *Visual Perception*. Academic, New York (1970)
6. Benary, W.: Beobachtungen zu einem Experiment über Helligkeitskontrast. *Psychol. Forsch.* **5**, 131–142 (1924)
7. Agostini, T., Galmonte, A.: Perceptual organization overcomes the effects of local surround in determining simultaneous lightness contrast. *Psychol. Sci.* **13**, 88–92 (2002)
8. White, M.: A new effect of pattern on perceived lightness. *Perception*. **8**(4), 413–416 (1979)
9. Knill, D.C., Kersten, D.: Apparent surface curvature affects lightness perception. *Nature*. **351**, 228–230 (1991)
10. Adelson, E.H.: Perceptual organization and the judgment of brightness. *Science*. **262**, 2042–2044 (1993)
11. Agostini, T., Proffitt, D.R.: Perceptual organization evokes simultaneous lightness contrast. *Perception*. **22**(3), 263–272 (1993)
12. Lotto, R.B., Purves, D.: The effects of color on brightness. *Nat. Neurosci.* **2**(1), 1010–1014 (1999)
13. Wertheimer, M.: Untersuchungen zur lehre von der Gestalt. *Psychol. Forsch.* **4**, 301–350 (1923)
14. Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., Economou, E.: An anchoring theory of lightness perception. *Psychol. Rev.* **106**, 795–834 (1999)
15. Kingdom, F., Moulden, B.: A multi-channel approach to brightness coding. *Vis. Res.* **32**, 1565–1582 (1992)
16. Blakeslee, B., McCourt, M.E.: A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vis. Res.* **44**, 2483–2503 (2004)
17. Todorović, D.: Lightness, illumination, and gradients. *Spat. Vis.* **19**(2–4), 219–261 (2006)
18. Robinson, A.E., Hammonb, P.S., de Sa, V.R.: Explaining brightness illusions using spatial filtering and local response normalization. *Vis. Res.* **47**(12), 1631–1644 (2007)
19. Galmonte, A., Soranzo, A., Rudd, M.E., Agostini, T.: The phantom illusion. *Vis. Res.* **117**, 49–58 (2015)
20. Gilchrist, A.: *Seeing Black and White*. Oxford psychology series. Oxford University Press (2006)
21. Eagleman, D.M., Jacobson, J.E., Sejnowski, T.J.: Perceived luminance depends on temporal context. *Nature*. **428**, 854–856 (2004)
22. Gilchrist, A.: The perception of surface blacks and whites. *Sci. Am.* **24**(3), 88–97 (1979)
23. Gilchrist, A.L.: Lightness contrast and failures of lightness constancy: a common explanation. *Percept. Psychophys.* **43**(5), 415–424 (1988)
24. Schirillo, J.: Surround articulation. I. Brightness judgments. *J. Opt. Soc. Am. A.* **16**, 793–803 (1999)
25. Soranzo, A., Agostini, T.: Does perceptual belongingness affect lightness constancy? *Perception*. **35**, 185–192 (2006)
26. Soranzo, A., Agostini, T.: Photometric, geometric and perceptual factors in illumination-independent lightness constancy. *Percept. Psychophys.* **68**(1), 102–113 (2006)
27. Soranzo, A., Gilchrist, A.: Layer and framework theories of lightness. *Atten. Percept. Psychophys.* **81**(5), 1179–1188 (2019). <https://doi.org/10.3758/s13414-019-01736-1>
28. Soranzo, A., Galmonte, A., Agostini, T.: Lightness constancy: ratio invariance and luminance profile. *Atten. Percept. Psychophys.* **71**(3), 463–470 (2009)
29. Soranzo, A., Galmonte, A., Agostini, T.: The luminance misattribution in lightness perception. *Psihologija*. **43**(1), 33–45 (2010)
30. Soranzo, A., Lugin, J.-L., Wilson, C.: The effects of belongingness on the SLC phenomenon: a VR study. *Vis. Res.* **86**, 97–106 (2013)

Sky Glow

- [Light Pollution](#)

Sleep Model

- [Sleep, Alertness, and Light](#)

Sleep, Alertness, and Light

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

[Sleep model](#)

Definition

Nighttime sleep and daytime alertness and performance are affected by daytime light through a circadian route and through a direct photobiological route.

Sleep, Health, Alertness, and Performance

Just as air, food, and water, sleep is a prime requirement for life. It saves bodily energy by slowing down many processes, and it has a restorative function. It is well established that insufficient sleep and sleep of low quality relates to all-cause mortality [1–4]. Health problems such as obesity, diabetes, cardiovascular diseases, and depressive disorders can be linked with insufficient sleep or bad sleep quality [5–7]. Sleep also plays a role in storing new information into our long-term memory [8]. The restorative function of sleep is not only fundamental for our health; it is also needed for adequate alertness and performance during the time that we don't sleep.

Model of How Sleep Is Regulated

In the early 1980s of the last century, cooperation between the universities of Zurich, Basel (Switzerland), and Groningen (the Netherlands) has resulted in a classical model of how sleep is regulated [9, 10]. It is still today the prevalent one [11]. The model is based on an interaction of two

different processes: a homeostatic and a circadian one. Homeostasis is a process that opposes changes. The homeostatic process in sleep-wake regulation increases sleep pressure (fatigue) after waking up and decreasing it during sleep. It is referred to as process S. Sleep pressure is behavior-dependent, namely, of the time being awake and of the previous sleep history. Sleep can occur when the circadian process (called process C) provides conditions for sleep. Process C is not behavior-dependent but dependent on the state of the body regarding, for example, melatonin and cortisol levels and body temperature. The amount and timing of light and dark influence circadian processes and thus also the sleep process C.

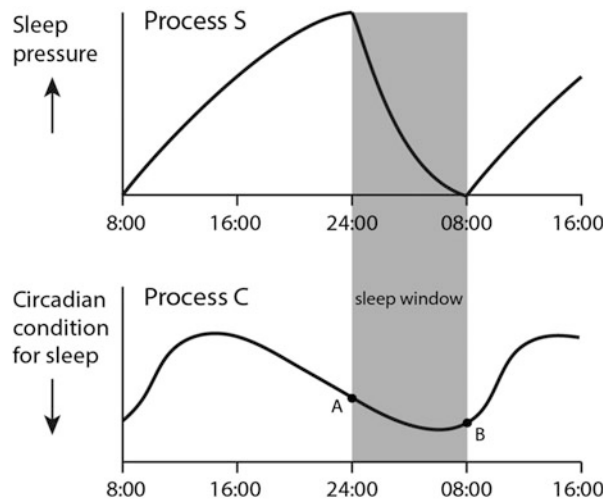
Figure 1 (top) shows an example of the typical exponential increasing sleep pressure after waking up together with the steeper, and again exponential, declining sleep pressure during sleep. Figure 1 (bottom) shows an example of the circadian curve where point A represents a situation where the conditions are such that sleep is possible and, further in time, point B where the conditions make sleep again difficult. The time frame between A and B represents the sleep window.

The model is a simplification of complicated behavior and bodily processes of which still much is unknown. A physiological confirmation of the homeostatic process has not yet been identified. Nevertheless, the model permits for the prediction of effects of sleep-wake scenarios and ratings of sleepiness, alertness, and performance [12, 13].

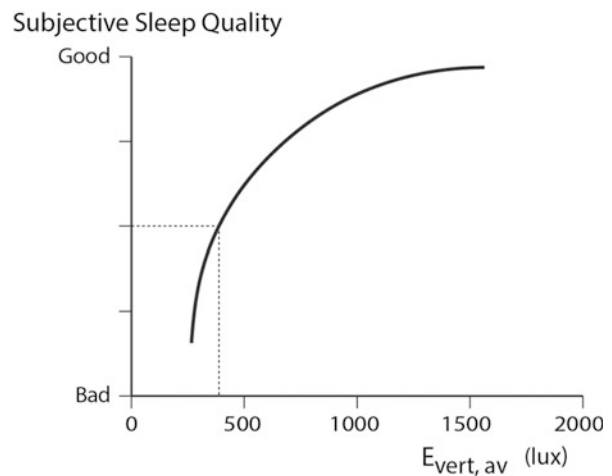
Daytime Light and the Circadian Process

Because the free-running rhythm of the circadian process, controlled by our biological clock (suprachiasmatic nucleus), is with most people different from the 24-h solar day, external cues which have an exact 24-h rhythm have to synchronize our biological clock with the 24-h rhythm. Under normal situations, the natural 24-h dark-light rhythm synchronizes the free-running rhythm with the 24-h rhythm. In chronobiology, the word “zeitgeber” has become an usual expression, also in English-language literature, for this “time-giver” effect of the transition of

Sleep, Alertness, and Light, Fig. 1 Two-process model of sleep regulation. Homeostatic process S drives sleep pressure (increasing after awakening and decreasing during sleep). Process C, the circadian process, provides the bodily conditions for sleep between points A and B: the sleep window



Sleep, Alertness, and Light, Fig. 2 Subjective nighttime sleep quality as a function of the average daytime vertical illuminance during work hours at the working place. Electrical lighting used 3000 and 4000 K fluorescent tubes [21]



dark and light. Especially morning light is a powerful zeitgeber.

Sufficient light during daytime and good darkness during nighttime is, therefore, a prerequisite for a sufficient nighttime sleep of good quality. A recent study with 20 test persons showed that light exposure measured throughout the 24 h preceding the sleep period correlates with the objective sleep quality [14]. This study made, for the first time, use of a novel, mobile PSG unit. It was combined with an action watch with a light sensor. Light levels varied roughly between 10 lux and 30.000 lux. The results of this investigation suggest not only a fundamental influence of light exposure on the circadian aspect of sleep but also an influence of light exposure on the

homeostatic sleep pressure aspect. As the authors of this study state: “It could change the current understanding of the regulation of sleep.” A clear relationship was found between daytime lighting and subsequent, objectively measured, nighttime sleep quality.

Light Level

Different studies carried out under interior daytime-lighting conditions in working interiors have shown that lighting levels that fulfil the requirements for visual performance and satisfaction do often not fulfil the requirements for sleep [15–20]. A study of Aries [21] illustrates this convincingly. She studied the relationship between vertical illuminance at the eye during

daytime and the subsequent nighttime sleep quality of 42 office workers in April and May in the Netherlands. She recorded the average vertical illuminance, from daylight and electrical light, during daytime working hours at the workstations of these office workers, spread over ten different office buildings. She determined the subjective sleep quality of each test person over the test period using the Groningen Sleep Quality Scale

Sleep, Alertness, and Light, Table 1 Illuminances at the eye corresponding to a circadian stimulus value CS = 0.3 for different light [20]

Light source	Illuminance (lux) corresponding to CS = 0.3
Daylight D65	175
White phosphor LED 6500 K	170
White phosphor LED 3500 K	400

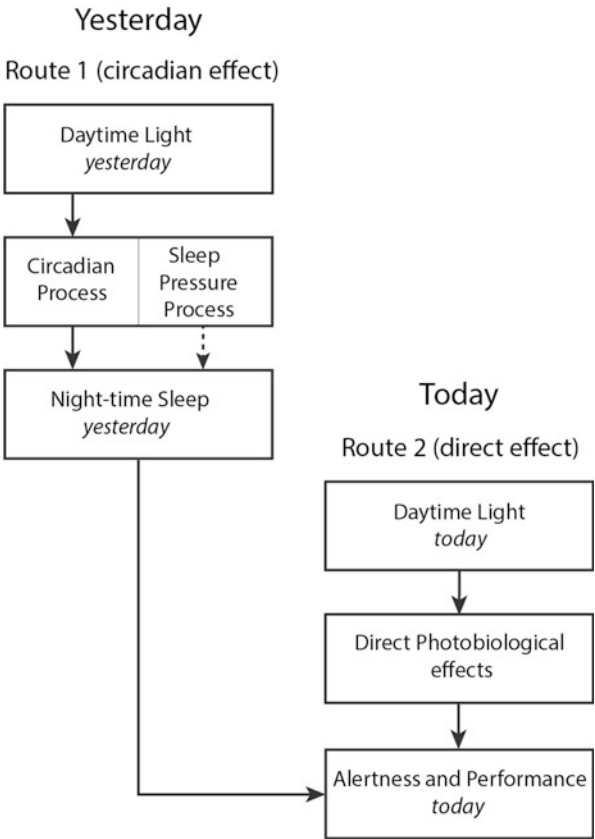
Sleep, Alertness, and Light, Fig. 3 The two routes through which daytime light may influence daytime alertness and performance

questionnaire. The results, given in Fig. 2, show that lighting during daytime correlates positively with nighttime sleep quality.

Figure 2 shows that the midpoint of the sleep quality scale corresponds to a value of 400 lux vertical illuminance: the bare minimum required for acceptable sleep quality. With the kind of light distributions, layout of luminaires typically used for office lighting, and correlated color temperatures between 3000 K and 4000 K, the average horizontal illuminance at working plane level corresponds to roughly three times the amount of vertical illuminance at eye height. A vertical illuminance of 400 lux thus corresponds to some 1200 lux horizontally, distinctly more than the values required today in most lighting standards for interior lighting for workplaces.

Light Spectrum

The studies discussed so far used the photopic illuminance as a parameter. Therefore, these



studies do not include the effects of the spectrum of the light. Figueiro [20] used the measure circadian stimulus (CS) as a parameter that includes the effect of the spectrum (see also the entry “Non-Visual Lighting Effects and Their Impact on Health and Well-Being” in this encyclopedia). For guidance of the practical meaning of a CS value of 0.3, the minimum required according to Figueiro, Table 1 gives the corresponding illuminance values at the eye for different light sources. With the light sources of today in interiors used for work (correlated color temperature around 3000–4000 K), the minimum requirement for morning light is at least 400 lux (at the eye). With light sources with CCT values of around 6500 K, this value can be considerably lower, down to slightly less than 50%.

Light, Sleep, Alertness, and Performance

Light affects sleepiness, alertness, and performance during daytime through two different routes (Fig. 3). Route 1 represents the circadian route and route 2 the direct photobiological one. Route 1 starts at the moment of daytime light on the previous day (“yesterday”). Daytime alertness and performance will obviously be better after a night with sufficient sleep of good quality. Daytime light of yesterday affects the nighttime sleep quality of yesterday. Therefore, the daytime lighting of yesterday influences sleepiness, alertness, and performance of today. Route 1 concerns long-term effects of light and lighting on circadian rhythms. Light received up to a few days before can cause these effects. Light can, however, also have short-term, direct effects within a short time after being exposed to light. These effects are called “direct photobiological effects.” These effects usually disappear shortly after the light exposure stops. An important example of a direct photobiological effect is light, both at day and night, acutely decreasing sleepiness and improving alertness [22–24]. The increase of alertness, in turn, has positive consequences for performance. Route 2, the direct photobiological route, concerns daytime light that invokes an acute

activating effect because of direct photobiological processes.

Cross-References

- [Circadian Rhythms](#)
- [Human-Centric Lighting](#)
- [Nonvisual Lighting Effects and Their Impact on Health and Well-Being](#)

References

1. Cappuccio, F.P., D’Elia, L., Strazzullo, P., Miller, M.A.: Sleep duration and all-cause mortality: A systematic review and meta-analysis of prospective studies. *Sleep*. **33**(5), 585–592 (2010)
2. Dijk, D.-J.: Sleep and health: Beyond sleep duration and sleepiness? *J. Sleep Res.* **21**, 355–356 (2012)
3. Odds, W. (ed.): *Sleep, Circadian Rhythms, and Metabolism; The Rhythm of Life*. Apple Academic Press, Inc., Oakville (2015)
4. Yin, J., Jin, X., Shan, Z., Li, S., Huang, H., Li, P., Peng, X., Peng, Z., Yu, K., Bao, W., Yang, W., Chen, X., Liu, L.: Relationship of sleep duration with all-cause mortality and cardiovascular events: A systematic review and dose-response meta-analysis of prospective cohort studies. *J. Am. Heart Assoc.* **6**(9), e005947 (2017)
5. Cappuccio, F.P., Cooper, D., D’Elia, L., Strazzullo, P., Miller, M.A.: Sleep duration predicts cardiovascular outcomes: A systematic review and meta-analysis of prospective studies. *Eur. Heart J.* **32**, 1484–1492 (2011)
6. Costa, I.C., Carvalho, H.N., Fernandes, L.: Aging, circadian rhythms and depressive disorders: A review. *Am. J. Neurodegener. Dis.* **2**(4), 228–246 (2013)
7. Kim, Y., Wilkens, L.R., Schembre, S.M., Henderson, B.E., Kolonel, L.N., Goodman, M.T.: Insufficient and excessive amounts of sleep increase the risk of premature death from cardiovascular and other diseases: The multi-ethnic cohort study. *Prev. Med.* **57**, 377–385 (2013)
8. Rasch, B., Born, J.: About sleep’s role in memory. *Physiol. Rev.* **93**, 681–766 (2013)
9. Borbély, A.A., Wirz-Justice, A.: Sleep, sleep deprivation and depression. A hypothesis derived from a model of sleep regulation. *Hum. Neurobiol.* **1**, 205–210 (1982)
10. Daan, S., Beersma, D.G., Borbély, A.A.: Timing of human sleep: Recovery process gated by a circadian pacemaker. *Am. J. Physiol.* **246**, R161–R178 (1984)
11. Borbély, A.A., Daan, S., Wirz-Justice, A., DeBoer, T.: The two-process model of sleep regulation: A reappraisal. *J. Sleep Res.* **25**, 131–143 (2016)

12. Åkerstedt, T., Folkard, S.: Predicting duration of sleep from the three process model of regulation of alertness. *Occup. Environ. Med.* **53**(2), 136–141 (1996)
13. Beersma, D.G.M., Gordijn, M.C.M.: Circadian control of sleep-wake cycle. *Physiol. Behav.* **90**, 190–195 (2007)
14. Wams, E.J., Woelders, T., Marring, I., van Rosmalen, L., DGM, B., MCM, G., Hut, R.A.: Linking light exposure and subsequent sleep: A field polysomnography study in humans. *Sleep*, zsx165 (2017)
15. Aries, M.: Human lighting demands, healthy lighting in an office environment. PhD thesis, University of Technology Eindhoven, The Netherlands (2005)
16. Viola, A.U., James, L.M., Schlangen, L.J.M., Dijk, D.-J.: Blue-rich white light in the workplace improves self-reported alertness, performance and sleep quality. *Scand. J. Work Environ. Health.* **34**, 297–306 (2008)
17. Hubalek, S., Brink, M., Schierz, C.: Office workers' daily exposure to light and its influence on sleep quality and mood. *Light. Res. Technol.* **42**, 33–50 (2010)
18. Boubekrie, M., Cheung, I.N., Reid, J.R., Wang, C.-H., Zee, P.C.: Impact of windows and daylight exposures on overall health and sleep quality of office workers: A case-control pilot study. *J. Clin. Sleep Med.* **10**(6), 603–611 (2014)
19. Canazei, M., Dehoff, P., Staggl, S., Pohl, W.: Effects of dynamic ambient lighting on female permanent morning shift workers. *Light. Res. Technol.* **46**, 140–156 (2014)
20. Figueiro, M.G., Steverson, B., Heerwagen, J., Kampschroer, K., Hunter, C.M., Gonzales, K., Plitnick, B., Rea, M.S.: The impact of daytime light exposures on sleep and mood in office workers. *Sleep Health.* **3**, 204–215 (2017)
21. Aries, M.: Human lighting demands, healthy lighting in an office environment. Public presentation of PhD thesis, University of Technology Eindhoven, The Netherlands (2005)
22. Phipps-Nelson, J., Redman, J.R., Dijk, D.-J., Rajaratman, S.M.W.: daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance. *Sleep.* **26**, 695–700 (2003)
23. Vandewalle, G., Balteau, E., Phillips, C., Degueldre, C., Moreau, V., Sterpenich, V., Albouy, G., Darsaud, A., Desseilles, M., Dang-Vu, T.T., Peigneux, P., Luxen, A., Dijk, D.-J., Maquet, P.: Daytime light exposure dynamically enhances brain responses. *Curr. Biol.* **16**, 1616–1621 (2006)
24. Cajochen, C.: Alerting effects of light. *Sleep Med. Rev.* **11**, 453–464 (2007)

Sodium Gas Discharge Lamps

- [High- and Low-Pressure Sodium Lamp](#)

Soft Colors

- [Pastel Colors](#)

Solid-State Light Source

- [Light-Emitting Diode, LED](#)
 ► [Light-Emitting Diode, OLED](#)

Sparkle

- [Texture Measurement, Modeling, and Computer Graphics](#)

Spectral Imaging

- [Hyperspectral Imaging](#)

Spectral Luminous Efficiency

János Schanda
 Veszprém, Hungary

Synonyms

[Eye sensitivity functions](#); [Luminous efficiency function](#); [Relative visibility function](#); $V(\lambda)$, $V'(\lambda)$

Definition

Spectral luminous efficiency (of a monochromatic radiation of wavelength λ) [$V(\lambda)$ for photopic vision; $V'(\lambda)$ for scotopic vision] (see CIE e-ILV [1]).

János Schanda: Deceased

Ratio of the radiant flux at wavelength λ_m to that at wavelength λ such that both produce equally intense luminous sensations under specified photometric conditions and λ_m is chosen so that the maximum value of this ratio is equal to 1.

Unit: 1

Overview

By the end of the nineteenth century, photometry achieved a level where it became necessary to describe the light perception efficiency of radiation of different wavelengths. Several methods were used to determine the spectral luminous efficiency function (called also spectral visibility). The two most important ones were direct comparison and the flicker method (based on the fact that if two lights are seen alternatively with increasing frequency, first the sensation of hue difference is lost and only later the sensation of intensity difference; choosing a frequency when color sensation is already lost but the intensity flicker is still perceived (in the order of 15 Hz), one can equate the “intensity” of the two lights by minimizing the flicker). Based on many investigations, CIE agreed to a table showing the relative visibility of lights of different wavelengths compared to the light with highest visibility in 1924 [2].

Spectral luminous efficiency $V(\lambda)$ for foveal vision at photopic levels: The present standard [3] is based on the table recommended in 1924, practically only interpolating and extrapolating the values with a very minor smoothing made. This $V(\lambda)$ function is the basis of present-day photometry, and due to the fact that the CIE 1931 colorimetry was based on photometry, the $2^\circ \bar{y}(\lambda)$ function was chosen to be equal to the $V(\lambda)$ function. The $V(\lambda)$ function is – more or less – valid (see later) at high luminance levels (accordingly, the $V(\lambda)$ function should be used above 5 cd/m² luminance, and it should mimic the spectral sensitivity of one of the cone signal channels (composed from the L and M cone signals) of human vision for foveal vision (part of the human retina where the spatial sensitivity (visual acuity) is highest.

This part of the retina is covered by a yellowish layer (macula lutea) that functions as a color filter. At larger eccentricity angles, where the macula lutea does not cover the fovea anymore, the eye sensitivity is different, see $V_{10}(\lambda)$ later).

Spectral luminous efficiency $V'(\lambda)$ for foveal vision at scotopic levels: At low light levels, the rods take over the role of vision; this is the range of scotopic vision. The spectral sensitivity of the rods differs from that of the cones. The spectral luminous efficiency for scotopic vision was internationally agreed in 1951 [5] and was standardized in 2004 [4]; it should be used if the luminance is smaller than 0.005 cd/m².

Mesopic spectral luminous efficiency $V_{mes}(\lambda)$: Between scotopic and photopic luminance levels, both rods and cones contribute to the spectral luminous efficiency. CIE published in 2010 a method to calculate the luminance level-dependent spectral luminous efficiency function in this mesopic range [4].

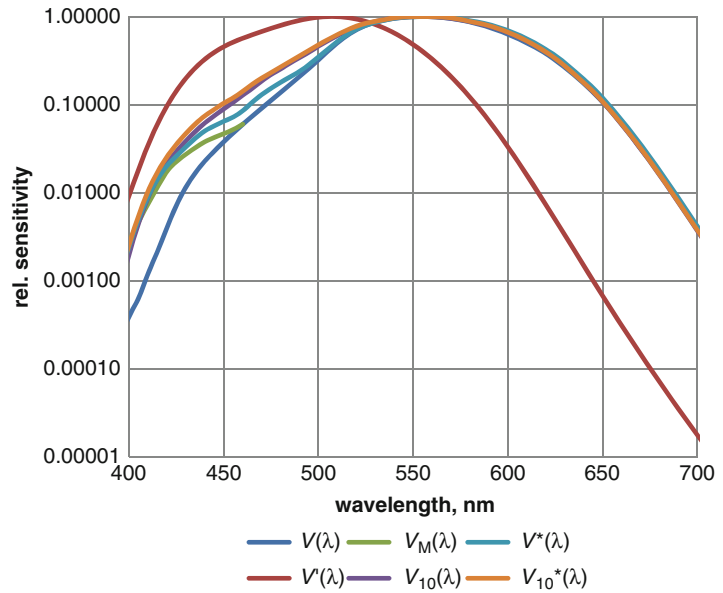
Some Further Visibility Functions

Modified 2° spectral luminous efficiency: It turned out quite early that the values of the 2° spectral luminous efficiency function are lower as would correspond to average human vision, but it took till 1990 that CIE adopted the $V_M(\lambda)$ function [6] and recommended it for applications in visual sciences. For further details on how this function has been derived and what forerunners have been published, see above publication.

Spectral luminous efficiency for large visual fields: There was no internationally accepted spectral luminous efficiency function till 2005, when CIE adopted the $V_{10}(\lambda)$ function based on the 10° standard colorimetric observer's $\bar{y}_{10}(\lambda)$ function [7].

Cone fundamental-based spectral luminous efficiency functions: Recent investigations have shown that in human vision, one of the visual channels provides a luminance-like signal to the brain. Based on detailed investigations,

Spectral Luminous Efficiency, Fig. 1 Some more important spectral luminous efficiency functions



CIE determined the average cone fundamental spectral response functions (spectral responsivities at human eye cornea level) and recommended color matching functions that correspond better to average human color perception as the CIE 2° and 10° standard observer functions [8].

Based on the above, Stockman and coworkers determined an average spectral luminous efficiency function for daylight adaptation [9] taking world population genotype distribution into consideration (the genotype influences the cone fundamental spectral sensitivity $V^*(\lambda)$).

In the following figure, we summarize the different spectral luminous efficiency functions in logarithmic representation. As can be seen, the traditional – and still official – photometric curve is way too low beyond 450 nm, having as a consequence erroneous photometric evaluation of blue LED lights. The $V_M(\lambda)$ curve makes some correction to this discrepancy. The $V^*(\lambda)$ curve is the most recent modification [9] based on cone fundamental signals, average population, and daylight adaptation (earlier $V(\lambda)$ functions have not taken any adaptation into consideration). The $V_{10}(\lambda)$ and $V_{10}^*(\lambda)$ curves differ from each other only very slightly, while the $V'(\lambda)$ function –

as it represents the input from a quite different type of light-sensitive cells – shows markedly different character (Fig. 1).

One has to emphasize, however, that up to now (2013), all practical photometric calculations are based on the CIE 1924 $V(\lambda)$ function and the Meter Convention has endorsed only the $V(\lambda)$ and the $V'(\lambda)$ functions. All other visibility functions have been recommended only for trial and some eye research applications. With the increasing use of LED-based light sources that contain a pronounced spectral maximum in the blue part of the spectrum, prediction errors using the $V(\lambda)$ function become apparent, and the use of more modern spectral luminous efficiency functions might become necessary.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Special Metamerism Index: Change in Observer](#)

References

1. CIE e-ILV item 17–1222 in <http://eilm.cie.co.at/term/1222>
2. CIE Compte Rendu 6e session, p. 67
3. CIE Photometry – The CIE System of Physical Photometry CIE S 010/E:2004/ ISO 23539:2005(E)
4. CIE Recommended System for Mesopic Photometry based on Visual Performance, CIE 191:2010
5. CIE Compte Rendu 12e session, vol. 3, p. 37
6. CIE 1988 Modified 2° Spectral Luminous Efficiency Function for Photopic Vision”, $VM(\lambda)$, CIE 86–1990
7. CIE 10° photopic photometric observer”, $V_{10}(\lambda)$, CIE 165:2005
8. CIE Fundamental Chromaticity Diagram with Physiological Axes – Part 1, CIE 170-1:2006, and draft of Part 2
9. Sharpe, L.T., Stockman, A., Jagla, W., Jägle, H.: A luminous efficiency function, $V^*(\lambda)$, for daylight adaptation. *J. Vis.* **5**, 948–968 (2005)

Spectral Power Distribution

Yi-Chen Chuang

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Spectral power distribution curves](#)

Definition

Spectral power distribution (SPD) is a function of wavelength and describes the amount of optical radiation across part of a spectrum. SPD usually describes optical radiation emitted by a light source, but generally, SPD can be applied to describe optical radiation reflected by an object, or transmitted from a transparent/semi-transparent object.

According to the definition of the Illuminating Engineering Society of North America (IESNA), SPD is “a pictorial representation of the radiant power emitted by a light source at each wavelength or band of wavelengths in the visible region of the electromagnetic spectrum.” Ideally,

SPD should be expressed in units of W/nm. However, in some of the literature, SPD is a general term used to describe any radiometric or photometric quantity as a function of wavelength (with corresponding units). In many conventional cases, SPD is referred to as radiant power per unit area as a function of wavelength, and this should be correctly termed spectral irradiance distribution (or radiance exitance distribution, depending on the context) expressed in units of W/m²/nm. Although based on the definition of the IESNA, SPD is limited to visible light. SPD is generally applicable to ultraviolet radiation, visible light, and infrared radiation. In color applications, SPD in the visible light region (380–780) nm is the main focus of interest, but SPD in the ultraviolet region is also important because it influences the color perception of fluorescent materials.

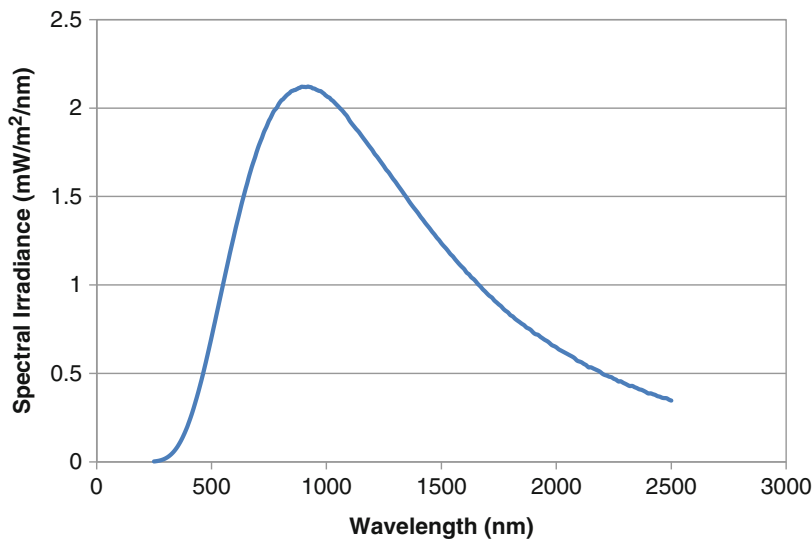
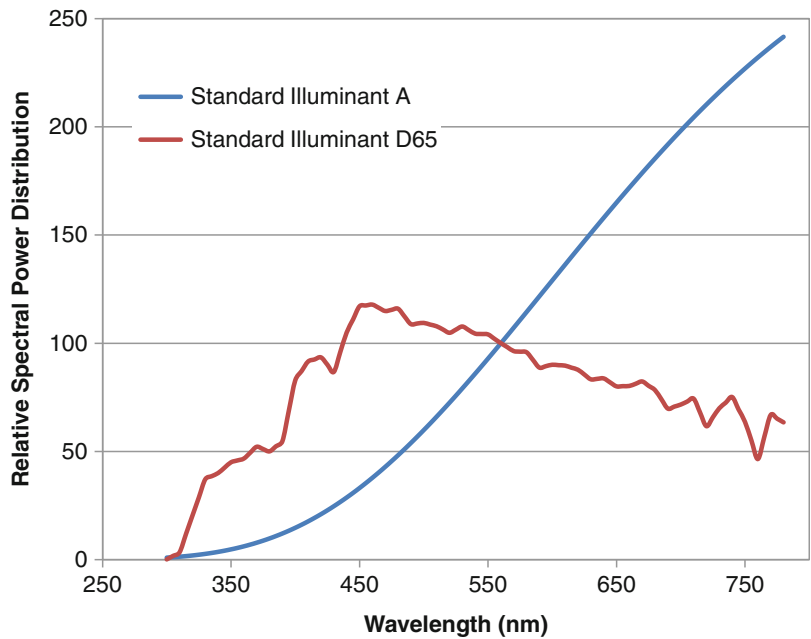
As the overall power levels of light sources can vary over many orders of magnitude, SPDs are often normalized for easy comparisons of color properties. Such SPDs are referred to as relative SPDs (dimensionless). A relative SPD is usually normalized so that it has a maximum value of 100 (or 1).

Overview

Spectral power distribution is a function of wavelength and describes the amount of optical radiation within a particular range of a spectrum [1–4]. SPD is used to describe not only light sources but also illuminants. The difference between light source and illuminant is that a light source is an actual physical entity, whereas an illuminant may only be a set of values recording its SPD. For example, International Commission on Illumination (CIE) Standard Illuminant A is the SPD of CIE Source A (approximated by an incandescent, tungsten filament light source). Many CIE Standard Illuminants do not exist, that is, there is no actual light source that can realize the SPD defined for such an illuminant. For example, the CIE D series of illuminants represents various phases of daylight, but no source is available for the D illuminants [1, 5]. Figure 1 shows the SPDs

Spectral Power Distribution,

Fig. 1 Spectral power distribution of International Commission of Illumination (CIE) Standard Illuminants A and D65



Spectral Power Distribution, Fig. 2 Spectral power distribution of a 1000 W tungsten lamp

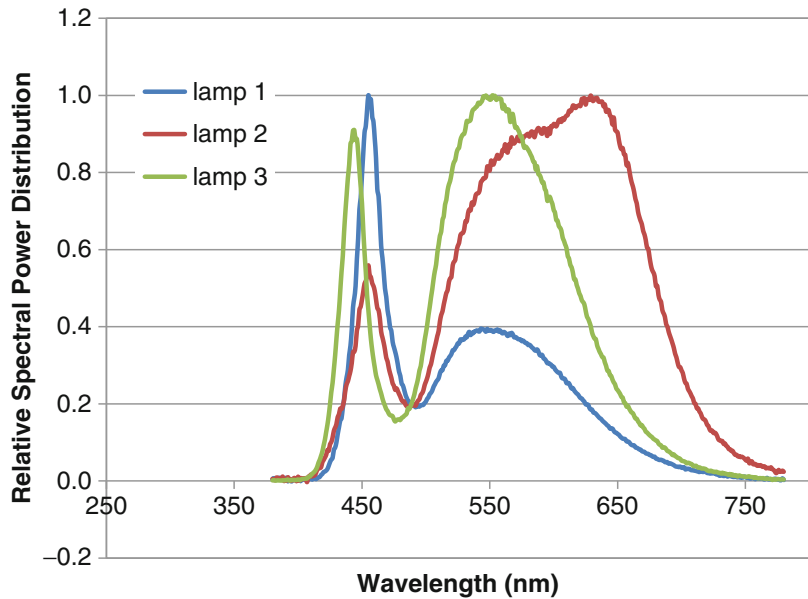
of CIE Standard Illuminant A and Standard Illuminant D65 from 300 to 780 nm, as examples of SPD [6].

In color science, SPD describes the color characteristics of a light source as a function of wavelength over the visible region. SPD is also often used to describe the color characteristics of an

object because optical properties such as transmittance, reflectance, and absorbance in addition to the detector response are typically dependent on the incident wavelength [7]. Knowledge of SPD contains all the basic physical data about the light and serves as the starting point for quantitative analyses of color. However, SPD cannot provide

Spectral Power Distribution,

Fig. 3 Sample spectral power distribution of white light emitting diodes



the human perception of objects [8]. Note that human perception is based on three components: SPD of light, reflectance of objects, and the color-matching function of human observers.

The SPD of light sources can be measured using a monochromator with optical detectors, or a spectroradiometer. An example of experimentally measured spectral irradiance distribution of a 1000 W tungsten lamp representing SPD is shown in Fig. 2. SPDs of other light sources can be found in a lamp manufacturer's manual, for example, or SPD simulators on-line for educational purposes [8].

By engineering SPDs of different light sources, it is possible to obtain the desired color property of light. These technologies have been applied to solid state lighting, such as white light emitting diodes (LEDs). Samples of SPDs of white LEDs are shown in Fig. 3. The “apparent white” of white LEDs is a “mixture” of different colors of light, where blue light is essential, as shown in Fig. 3. There are many ways to obtain precise color quality parameters, such as correlated color temperatures, illuminance, color rendering, luminous efficacy, etc. Note that different SPDs of light can result in similar apparent colors of light, but quite different color quality.

On the other hand, the apparent color of an object can be described by the SPD of the

reflected light from an illuminated surface and is the product of the percentage of reflectance of the surface and the SPD of the light that falls on the surface. Thus, an object may appear differently affected by different light illuminations.

Cross-References

- CIE Color Rendering Index
- CIE Standard Illuminants and Sources

References

1. Webster, J.G. (ed.): The Measurement, Instrumentation, and Sensors Handbook. Springer, Heidelberg (1999)
2. Peres, M.R.: The Focal Encyclopedia of Photography. Focal Press, Amsterdam (2007)
3. Fairchild, M.D.: Color Appearance Models. Wiley, Hoboken (2005)
4. Grum, F.C.: Optical Radiation Measurements, vol. 1. Academic, New York (1979)
5. CIE, Technical Report: Colorimetry, 3rd ed., Publication 15:2004, CIE Central Bureau, Vienna (2004)
6. Joint ISO/CIE Standard: ISO 10526:1999/CIE S005/E-1998 and CIE Standard Illuminants for Colorimetry
7. McCluney, W.R.: Introduction to Radiometry and Photometry. Artech House, Boston (1994)
8. <http://hyperphysics.phy-astr.gsu.edu/hbase/vision/spd.html> HyperPhysics (©C.R. Nave, 2012)

Spectral Power Distribution Curves

► Spectral Power Distribution

Spectrometer

Yi-Chen Chuang

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Spectroradiometer

Definition

A spectrometer is an optical instrument designed to measure the spectrum and stimulus of a light source or self-luminous object. The prefix “spectral” is an abbreviation of the phrase “spectral concentration of,” which is defined by the International Commission on Illumination (CIE) as the “quotient of the radiometric quantity taken over an infinitesimal range on either side of a given wavelength, by the range”.

Overview

Basic Principles

A spectrometer receives light, disperses the light, and converts the optical signals (or more generally, electromagnetic radiation signals) to electrical signals. It provides the radiometric quantities in narrow wavelength intervals. Because many sources have line structures, it is useful to sample the spectrum with narrow bandwidth and wavelength increments [1, 2]. In general, spectral irradiance varies from point to point on the surface. In practice, it is important to note how radiant flux varies with direction, the size of the solid angle subtended by the source at each point on the

surface, and the orientation of the surface. Given these considerations, it is often more prudent to use a stricter form of the formula, to account for these dependent relationships [3]. Spectral irradiance is the desired measurement for general spectroradiometry. In practice, the average of the spectral irradiance is measured, shown mathematically as the approximation:

$$E(\lambda) = \frac{\Delta\Phi}{\Delta A \Delta\lambda}$$

where Φ is the radiant flux of the source (SI unit: watt, W) within a wavelength interval $\Delta\lambda$ (SI unit: meter, m), incident on the surface area, A (SI unit: square meter, m²), and the calculated result E is the spectral irradiance. It is more useful to measure wavelength in nanometers; therefore, the submultiples of the SI units of spectral irradiance are often used, for example, $\mu\text{W}/(\text{cm}^2 \cdot \text{nm})$ [3].

A spectrometer is very similar to a spectrophotometer, another optical instrument used to measure an electromagnetic radiation spectrum. However, there are some differences. The name spectrometer is applied to instruments that operate over a very wide range of wavelengths to measure self-luminous objects. An optical spectrometer could include the UV–VIS–IR spectrum; spectrophotometers are usually used in spectral regions near the visible light spectrum. In general, the reflectance or transmittance is measured. For color measurements, spectrometers are used to measure the color of a light source, while spectrophotometers are used to measure the color of an object [2].

Applications

Spectrometers take the spectral radiance of measurements and spectral irradiance is typical. Based on the spectral data measured, the radiometric, photometric, and colorimetric quantities of light can be calculated, providing a complete description of the color of the source [3]. Thus, the spectrometer is an important tool in the color industry as the basis for calibration. Quantities commonly used in the color industry include CIE tristimulus values, CIE chromaticity

coordinates (x, y), correlated color temperature (CCT), color rendering index (CRI), etc. [2].

An automatic scanning spectrometer includes a mechanism to change the wavelength selected by the monochromator and to record the resulting changes in the measured quantity as a function of the wavelength. Many commercial spectrometers have included built-in post-calculation-functions for easy usage of the spectrometer for various applications. The latter is the basic configuration of a spectrometer, which usually includes a dispersing device and a detecting device (i.e. an optical detector or alternative).

This configuration allows the simultaneous analysis of colors. Such an instrument can record a spectral function without mechanical scanning, which may be corrected in terms of resolution or sensitivity, for instance. The absorption spectrophotometer measures the absorption of light by a sample as a function of wavelength.

References

1. Spectroradiometer: http://en.wikipedia.org/wiki/Spectroradiometer#cite_note-3
2. Janos, S.: Colorimetry: Understanding the CIE System. Wiley, Hoboken (2007)
3. Illuminating Engineering Society (IES): Lighting measurements: a basic approach to understanding the principals of lighting science. Handour. Comp. Nlena and Associates. Retrieved from <http://www.ies.org/pdf/education/IES-Color-3-Webcast-Handout.pdf>
4. Stroebel, L.D., Zakia, R.D.: Focal Encyclopedia of Photography, 3rd edn, p. 115. Focal Press, London (1993). ISBN 0-240-51417-3

Spectroradiometer

► [Spectrometer](#)

Spectroreflectometer (for Reflectance Measurement)

► [Instrument: Spectrophotometer](#)

Sports Lighting

Gilles Page

Signify, Miribel, France

Definition

The purpose of sports lighting is to create the ideal conditions of visibility for those who experience sports facilities.

User Requirements

The great interest of the media in sports events has an important effect on the development of the sports themselves. This, together with the increased leisure time, has resulted in a booming number of active sportsmen and women all over the world, either taking part in the sports as a leisure activity or in amateur and professional competitions.

Consequently, numerous sports complexes have been developed in which all kinds of sports are practiced, from the general physical training sessions to professional activities. In the organization of sports activities and events, the quality of the facilities available is for the utmost importance for pleasure and in the outcome of the competition. Space, a good playing surface, heating/cooling, safety provisions, etc. are all of major importance.

One essential contribution to a successful sports venue, which is still not always recognized as such, is the standard and quality of the lighting system. This, of course, is not only for TV transmissions, where lighting levels of sufficient quality are an absolute necessity to assure an optimal television transmission, but also for the general practicing of the sports, guaranteeing, for example, a clear view of the components, the objects like the ball, and, of course, the indispensable marking of the competitive area itself.

The users of sports grounds can be distinguished according to their activities:

- The players
- The technical staff, referees, and team officials
- The spectators arriving, watching the game, and leaving the sports facility
- Both the television and film crews and the photographers recording the event

The players, referees, and officials must be able to see clearly all that is going on in the competition area so that they can produce their best possible performance and make accurate decisions.

Spectators should be able to follow the performance of the players and the action in an agreeable environment. The latter requirement means that they must be able to see not only the competition area but also its immediate surroundings. The lighting should also enable spectators to safely enter and leave the sports facility. With large groups this safety aspect is very important.

For television and film coverage, the lighting must be able to ensure high-quality color images, not only of the overall action but also close-ups of the players and spectators alike. Closeup images are important to convey the emotions and atmosphere in a stadium to viewers watching at home.

Sports Lighting Classes

As the competence level of athletes increases so does the speed of the action; the visual task becomes more difficult, requiring more light of a higher quality.

In general, five levels of sporting activity are recognized: international and national competition, regional competition, local competition, training, and recreational. These levels are related to the standards of play and the viewing distances of the spectators. These five levels do not all require the same quality of lighting. Lower standards are clearly acceptable for recreation than for national competitions. To cover all five activity levels, three lighting classes are defined:

Class I: Top-level competition such as national and international matches, which generally

Sports Lighting, Table 1 Lighting classes according to CEN [1]

Level of activity	Class I	Class II	Class III
International and national competition	•		
Regional competition	•	•	
Local competition	•	•	•
Training		•	•
Recreational			•

involve large spectator capacities with potentially long viewing distances. Top-level training may also be included in this class.

Class II: Mid-level competition such as regional or local club matches which generally involve medium-size spectator capacities with medium viewing distances. High-level training may also be included in this class.

Class III: Low-level competition such as local or small club matches which do not usually involve spectators. General training and recreation also come into this class (Table 1).

Lighting Criteria

Knowing the general user requirements, it is possible to determine the lighting criteria for each of the different level of activity. The purpose of this section therefore is to identify these lighting criteria and, wherever possible, to derive the lighting parameters of interest in each case.

Illuminance on a Horizontal Plane

The illuminated playing surface forms a major part of the field of view for the players, officials, and spectators. It is the illuminance on this horizontal plane, at ground level, commonly referred to as the horizontal illuminance (E_h), which chiefly serves to establish the adapted state of the eye, by creating a stable visual background against which players and objects will be seen. Because of this an adequate horizontal illuminance on the field is important.

To ensure safety of movement for the spectators when entering and leaving the stands or surrounds, adequate horizontal illuminance in these areas is also required.

Vertical Illuminance

Because it is important to follow the progress of the players and the ball which can be approximated as vertical surfaces, it is the illuminance on the vertical surfaces (E_v) which describes it best. However, experience shows that where the human observer is concerned, there is a good relationship between vertical and horizontal illuminance. Therefore, for all activities with the exception of television broadcasting, if the specified horizontal illuminance is provided and the design rules are followed, the vertical illuminance will be sufficient.

The scene illuminance, and more particularly the vertical illuminance, has a major influence on the quality of the final television or film picture. To guarantee an optimal view and identification of players from all directions, specified illuminances on vertical planes are required. For television or filming with fixed camera positions, it is sufficient to ensure that the illuminances on planes at right angles to the camera positions are adequate. In the case of an unrestricted choice of camera positions, the vertical illuminance on planes facing all four sides of the competition area should be considered.

Uniformity

Good illuminance uniformity is important in order to avoid both adaptation problems for players and spectators. If the uniformity is not adequate, there is a risk that an object or player details, like sponsor logos and emotions, will not be clearly seen at certain positions on the competition area.

Uniformity is expressed as the ratio of the lowest to the highest (E_{min}/E_{max}) illuminance and as the ratio of the lowest to the average ($E_{min}/E_{average}$) illuminance.

Even when the uniformity ratios as defined above are acceptable, changes in illuminance can be disturbing if they occur over a too short distance. This problem is most likely to arise

when panning a television camera. Therefore, the illuminance uniformity for TV/film coverage at a certain grid point has to be expressed as a percentage change from the average adjacent grid points. This is called the gradient.

Glare Restriction

Glare is a subjective factor, for which a practical evaluation system has been described for outdoor sports applications, by CIE (*CIE 112 glare evaluation system for use within outdoor sports and area lighting*), based on extensive field tests [2]. CIE 112 defines a so-called glare rating factor (GR) for which an assessment scale is given from 10 to 90. The lower the value of GR, the better glare situation is. A maximum GR value of 50 is generally specified in high-end sports projects. The validity of the system is restricted to viewing directions below eye level. The development of the glare rating GR assessment method has led to a reliable system that can be used to make a qualitative statement about the suitability of a lighting system before it is installed.

This assessment method is supposed to be used both for predicting degree of glare at the design stage and checking glare situation of installation, but in reality it is almost never used for the latter situation as it is quite impossible (very difficult) to measure reliable glare on site.

It is also a reason why it is important to refer to guidelines for the positioning of floodlights.

Modeling and Shadows

Modeling is the ability of the lighting to reveal form and texture. This “modeling” ability is particularly important to provide a pleasant overall impression of the athletes, objects, and spectators on and around the pitch.

An installation where light only comes from one direction will cause harsh shadows and poor modeling. The players will appear to be pushed into the background and will only be seen from the same direction as which the light comes. From all other directions they will be seen as dark silhouettes only.

Color Properties of Lamps

Color perception is important in most sports, and while some color distortion, attributable to the artificial lighting, is acceptable, there should not be problem being able to distinguish between different colors.

There are two important aspects of the color properties of lamps:

The color appearance of the light. This is the color impression of the total environment created by the lamp.

The color-rendering properties of the lamps used or CIE color-rendering index. This describes how faithfully a range of colors can be reproduced by a light source.

Both the color appearance and the color-rendering properties of lamps are completely dependent upon the spectral energy distribution of the light emitted. An indication of the color appearance of a lamp can be obtained from its correlated color temperature in degrees Kelvin indicated as T_k , which vary mainly between 2000 and 6500 k. The lower the color temperature, the “warmer” the color impression of the light is; the higher the color temperature, the “cooler” or more bluish the impression of the light.

The color-rendering properties of a light source can be indicated by its color-rendering index R_a expressed as numerical values from 0 to 100. A light source having a color-rendering index of 100 will represent all scene colors faithfully and can be compared with daylight.

Because the color rendering is based on the human eye response, a new metrics based on TV camera response, the TLCI, has been developed by the EBU [3]. It is gaining popularity by broadcasters as being a color metric specifically designed for their needs.

Lighting Recommendations

Taking the various lighting criteria discussed in the previous section as a basis, it is possible to draw up two sets of lighting recommendations

for sports venues: one for venues without TV coverage and one for TV coverage, both being derived from the recommendations of the European standard EN 12193 – Sports Lighting [1].

The first set, presented in Table 2, concerns the lighting needs of players, officials, and spectators and covers a wide range of sporting activities and various levels of play, respectively, for outdoor and indoor activities without television coverage.

It should be noted that different practices are used in the USA; the Illuminating Engineering Society (IES) publishes illuminance recommendations for sports and recreation in its lighting handbook [4] that differ from the ones defined in the EN 12193; the main differences are the following:

- Four lighting classes defined.
- Nearly all sports activities have both a horizontal illuminance and a vertical illuminance component.
- Illuminance criteria are based on visual ages of more than half the intended observers.

Therefore, the recommended lighting levels are generally higher than the ones defined in the EN 12193 norm.

The second set is about lighting needs for TV coverage. When considering the lighting for broadcasting, the ideal situation is for the lighting to achieve the best possible standard relative to the events that will be regularly televised and the use of the venue. EN 12193 considers two categories: minimum for acceptable broadcasting and best practice for major events; the lighting criteria corresponding to these two categories are presented in Table 3.

For venues used for professional sports, where TV transmissions happen, the vertical illuminance has a major influence on the quality of a final television picture. EN 12193 requirements call for minimum vertical illuminance at any point of the field of play; this is the minimum lighting level under which the camera can predictably operate and produce quality images of the athlete. Other publications from national or international sports federations exist and may provide slight

Sports Lighting, Table 2 Lighting recommendations for non-televised activities

	Outdoor venues						Indoor venues					
	Class I			Class II			Class III			Class I		
	Eh (lux)	U	Eh (lux)	Eh (lux)	U	Eh (lux)	Eh (lux)	U	Eh (lux)	Eh (lux)	U	Eh (lux)
American football	500	0.7	200	0.6	0.5	75	200	0.5	200/1000-2000	200/1000-2000	0.5/0.8	200/1000-2000
Archery/ target 25 m-50 m	200	0.5	200	0.5	0.5	200	200	0.5	200/1000-2000	200/1000-2000	0.5/0.8	200/1000-2000
Athletics	500	0.7	200	0.7	0.5	100	200	0.5	500	300	0.6	200
Badminton							750	0.7	750	500	0.7	300
Bandy	750	0.7	500	0.7	0.5	200	200	0.5				
Baseball	750/500	0.7/0.5	500/300	0.7/0.5	0.5/0.3	300/200	300/200	0.5/0.3				
Basketball	500	0.7	200	0.6	0.5	75	750	0.5	750	500	0.7	200
Beach volley	500	0.7	200	0.6	0.5	75	750	0.5	750	500	0.8	500
Billiards, snooker												
Bobsleigh, luge	300	0.7	200	0.5	0.4	50	50	0.4				
Boules, pétanque	200	0.7	100	0.7	0.5	50	300	0.5	300	200	0.7	200
Boxing							2000	0.8	2000	1000	0.8	500
Cricket (infield/outfield)	750/500	0.7/0.5	500/300	0.7/0.5	0.5/0.3	300/200	750	0.5/0.3	750	500	0.7	300
Curling							300	0.7	300	300	0.7	300
Darts oche/target							200/750		200/750	100/500		50/300
Equestrian	500	0.7	200	0.7	0.5	100	500	0.5	500	300	0.6	200
Fencing							750	0.7	750	500	0.7	300
Football	500	0.7	200	0.6	0.5	75	750	0.5	750	500	0.7	200
Golf (tee/fairway/green)					0.5/0.3/0.5	100/20/50						
Gymnastics							500	0.7	500	300	0.6	200
Handball	500	0.7	200	0.6	0.5	75	750	0.5	750	500	0.7	200
Hockey	500	0.7	300	0.7	0.5	200	750	0.5	750	500	0.7	300
Horse racing	200	0.6	100	0.4	0.2	50						
Ice hockey	750	0.7	500	0.7	0.5	200	750	0.5	750	500	0.7	300
Judo, karate, Wrestling							750	0.7	750	500	0.7	200

(continued)

Sports Lighting, Table 2 (continued)

	Outdoor venues				Indoor venues			
	Class I		Class II		Class III		Class I	
	Eh (lux)	U	Eh (lux)	U	Eh (lux)	U	Eh (lux)	U
Rugby	500	0.7	200	0.6	75	0.5		
Shooting	200	0.5	200	0.5	200	0.5	200	0.5
Skatepark areas	100	0.6	50	0.4	30	0.3		
Skiing, ski jumping	100	0.5	30	0.3	20	0.2		
Softball (infield/outfield)	750/ 500	0.7/ 0.5	500/ 300	0.7/ 0.5	200/100	0.5/0.3		
Speed skating	500	0.7	200	0.7	100	0.5	500	0.7
Squash							750	0.7
Swimming	500	0.7	300	0.7	200	0.5	500	0.7
Table tennis							750	0.7
Tennis, padel	500	0.7	300	0.7	200	0.6	750	0.7
Track cycling	500	0.7	300	0.7	100	0.5	750	0.7
Volleyball	500	0.7	200	0.6	75	0.5	750	0.7
Weight lifting							750	0.7

Color rendering should be >60 for class II and III (indoor and outdoor) and > 70 for class I outdoor and > 80 for class I indoor
For outdoor venues glare rating should be <50 for class I and II and < 55 for class III
For indoor venues glare rating should be <35 for class I and < 40 for class II and III
Eh average horizontal illuminance on the field of play
U uniformity of the horizontal illuminance = Emin/Eaverage of the horizontal illuminance on the field of play

Sports Lighting, Table 3 Lighting recommendations for televised activities (indoor and outdoor)

		Minimum requirements	Major events
Vertical illuminance	Emin (lux)	600	1400
	Emin/Eaverage	0.6	0.7
	Gradient (% per m)	< 6	< 5
Horizontal illuminance	Eaverage (lux)	0.5 to 2 times Ev average	0.5 to 1.5 times Ev average
	Emin/Emax	0.5	0.6
	Emin/Eaverage	0.7	0.8
	Gradient (% per m)	< 5	< 5
Light source	T(K)	5000–6000	5000–6000
	TLCI	> 60	> 80
	Flicker (%)	< 5	< 1
Glare	GR indoor/outdoor	< 40/< 45	< 35/< 40

differences in the recommendation values and requirements.

Before consulting these tables, the following points should be noted.

Horizontal Illuminance

The horizontal illuminance needed on the playing area for players, officials, and spectators depends on several factors like:

- The size of the playing object
- The reflectance and color of the playing object relative to the reflectance and color of the background surfaces
- The viewing distances
- The speed of the playing object
- Whether the sports facility is located indoors or outdoors

Levels of illuminance for outdoor facilities are usually lower than for the same sports when played indoors; this is due to the corresponding improvement in contrast and adaptation that occurs when viewing a lit object against a dark background such as the night sky.

The illuminances on a horizontal plane needed for competitions with TV coverage are not listed.

Such levels are secondary to the requirements expressed in terms of vertical illuminance.

To ensure the television picture has a well-balanced brightness, the ratio between the average vertical and horizontal illuminance should be as

closely matched as possible but should not exceed the ratio of 0.5–2 times. That is, the horizontal illuminance should not be less than half the vertical illuminance or greater than twice vertical illuminance.

Vertical Illuminance

In practice, the vertical illuminance required for players and spectators will usually automatically be obtained if the requirements regarding the horizontal illuminance are fulfilled.

Where television broadcasting is a requirement, it is necessary to provide an adequate vertical illuminance across the scene viewed by the camera. If the vertical illuminance is not sufficient, good quality broadcast pictures will not be possible.

Television cameras are not able to adapt to changes and fluctuations in lighting level as quickly as the human eye. Therefore, this limitation must be considered when designing lighting systems for televised events.

In most of the sports activities, television cameras will be viewing predominantly vertical objects between the ground and a height of 1.5 m above the ground. In that case, the illuminance should be specified as the illuminance on a vertical plane 1.5 m above the ground facing the direction of the television camera. In other cases, when the main action takes place at a different height, the illuminance should be specified where the main action takes place; for example, for diving the vertical illuminance should be

specified on a vertical plane along the diving fall and into the direction of the camera.

It is now common that many cameras are used, distributed around the arena to obtain closeup action shots from alongside each event area. So, it is important to ensure that the vertical illuminance towards each camera is sufficient.

The recommended illuminances (horizontal and vertical) listed in Tables 2 and 3 are therefore “maintained” values under which the illuminances should never fall during the whole period of operation of the installation.

Illuminance Uniformity

The uniformity requirements for both horizontal and vertical illuminance are more stringent for TV and filming than for the players and spectators, the human eye being less sensitive to nonuniformities than are cameras.

It is principally the horizontal uniformity that determines the brightness range of the overall scene, so this has to satisfy the more stringent requirements.

Color Properties of Lamps

In sports application where color recognition is an important factor especially in case of media coverage, it is essential to use high-color-rendering light sources. For non-televised events, then a lower color-rendering index not less than 60 and up to 80 is generally specified.

For high-end sports events with TV coverage, it is important that true color scenes are properly seen through the TV camera signal. EN 12193 recommends using light sources with a TLCI >80 for major events and at least 60 as a minimum requirement for other cases.

Except in case of TV and film coverage, the color temperature quoted in the recommendations is merely to ensure an acceptable ambiance.

In order to acquire natural color scenes, the camera needs to be white balanced. Because a camera system can only adapt to one color temperature at a time, lighting systems used in combination with daylight should have a color temperature close to that of daylight. In addition, the preferred photographic films for sports usage are daylight balanced to around 5500 K, so light

sources with a color temperature between 5000 and 6000 K are generally recommended in case of TV coverage.

Flicker Effect

A problem for super slow-motion cameras is 50 Hz flicker due to the phasing of the light when using high-intensity discharge lamps. Cameras perceive changes in light levels due to the uneven ratio between the camera scanning frequency and the alternating amplitude of artificial lights powered by mains frequency.

An AC lamp “flashes” at twice the mains frequency. Due to the speed of the flashes, they are imperceptible to the human eye, and the illusion is created that the lighting is continuous. The exposure time of a standard camera corresponds to twice the lighting period. Therefore, a standard camera is integrating each field with an equal amount of light. However, when shooting with a high-speed camera, the behavior of artificial lighting will be shown on the picture, exactly as it is occurring.

In triple-speed mode (150 Hz), the exposure time is one third of the exposure time in single-speed mode. The alternating light amplitude causes different video levels for the three fields. This effect is only visible during slow-motion replay. Since a few years, this problem can be solved by using high-intensity discharge lamps coupled with electronic gear or simply using LED luminaires with appropriate electronic drivers.

Glare Restriction

The visual performance is reduced by glare which shall therefore be limited.

In sports lighting applications, mounting height of the luminaire is a key factor influencing the quality of an installation. This parameter has a big impact on glare and spill light; so, it has to be carefully defined, in relation with the aiming pattern, in order to limit glare while complying with the required lighting level and uniformities.

Additionally, it is wise to avoid installing floodlights in the main viewing directions of the athletes.

Environmental Aspect

The control of stray light from outdoor lighting installations can also be disturbing to people outside the lighted area, for example, to traffic on adjacent roads and to inhabitants of houses in the neighborhood of the sports facility. In some cases, local authorities or municipalities have their own guidelines on such matters, and their advice should always be sought.

It should be noted that stray light falling outside the playing area cannot be effectively controlled by lowering the mounting heights of the floodlights. This only results in luminaires being aimed in a more vertical direction causing even more glare.

In large stadiums it is now a common practice for large video screens or scoreboard systems to be used. In these cases, care must be taken to ensure that luminaires are not positioned too close to these screens as this could cause glare for persons looking up at the screen.

Safety Lighting for Participants

Participant safety is ensured by the safe stopping of an event which might otherwise be dangerous to continue in the absence of lighting.

European Norm EN 12193 about sports lighting specifies safety lighting for some sports activity where the absence of lighting could create a big danger.

The safety lighting shall come on the instant the general lighting fails and last for at least the period specified.

Emergency Lighting

For the purpose of safety and orientation for the spectators, as well as anti-panic lighting in case of mains power failure or emergencies, it is recommended to refer to European Norm EN 1838 (Lighting applications – Emergency lighting) [5].

Continuation of a Sport

For continuation of a sports, the lighting level shall be at least the class III level specified for that sports (see Table 2).

Broadcasting Continuity

The delaying or cancelation of a high-end sports event due to the loss of lighting is unacceptable. Where this is likely to cause a problem, e.g., at professional games and/or where television broadcasting is involved, then provision must be made to give an alternative solution.

Especially for major events broadcasted on TV, it is mandatory that, in the event of power failure, the continuity of the television broadcast shall be guaranteed. The chosen solution has to fit the specification and be relevant for the competition or event to be held in the stadium.

The actual requirement for an alternative power supply and switching method should be specified by sport's governing bodies and their contractual obligations towards the host broadcaster.

The main points to consider are the following:

- Time delay between switching from one power source to another, including the restrike time of lamps
- Lighting level to ensure broadcasting continuity

The highest standard would run the lighting via uninterruptible power supply (UPS) system with an alternative power supply available to supply 100% of the system. This system will provide a complete system with no reduction in lighting quality.

For lower standards (lower continuity lighting level and/or time delay accepted), other alternatives are available depending on several factors like the quality of the local/national power network, investment costs, and maintenance costs.

In that case, solutions can be defined by coupling several supply sources (electricity network, running generators, backup generators).

Spectator Area Lighting

Much of the existing documents about sports lighting mainly specify lighting recommendation on the field of play without taking care of the spectator areas.

For the visual comfort of the spectators, the European norm EN 12193 recommends a minimum value of 10 lx in the spectator area.

In case of TV coverage, it could happen that the broadcaster would specify a lighting level in the spectator area as a ratio of the field of play lighting level in order to suit his production style.

Light Sources

High-intensity discharge lamps have been used for many years as almost the unique light source for sports lighting applications. Since a few years, LEDs are taking over from high-intensity lamps.

A few advantages of the LED versus the high-intensity discharge lamps are:

- The ability to ignite instantaneously whatever the conditions (see section “[Broadcasting Continuity](#)”)
- Easy to create low flicker thanks to electronic drivers, so eliminating the flicker effect in case of slow-motion camera use
- Instant switch on/switch off + individual control of every luminaire, giving the possibility to create light shows before or after the games or during breaks
- Longer lifetime allowing to reduce maintenance cost

Therefore, LED luminaires are more and more adopted for professional venues but also for non-televised event venues for new installations as well as for refurbished ones.

Cross-References

- [CIE Color Rendering Index](#)

References

1. EN 12193:2018: Light and Lighting – Sports Lighting. CEN European Committee for Standardization, Brussels (2018)
2. CIE 112: 1994: Glare Evaluation System for Use within Outdoor Sports and Area Lighting. CIE Central bureau, Vienna (1994)
3. EBU Tech 335: Method for the Assessment of the Colorimetric Properties of Luminaires: the Television Lighting Consistency Index (Tlci-2012) and the Television Luminaire Matching Factor

(Tlmf-2013). European Broadcast Union, Geneva (2017)

4. IES Lighting Handbook, 10th edn. Reference and Application. Illuminating Engineering Society of North America, New York (2011)
5. EN 1838:1999: Lighting Applications – Emergency Lighting. CEN European Committee for Standardization, Brussels (1999)

Stain

- [Pigment, Ceramic](#)

Standard Daylight Illuminants

- [Daylighting](#)

Standard Illuminating and Viewing Conditions

- [Standard Measurement Geometries](#)

Standard Lamp

Cheng-Hsien Chen

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Reference standard lamp](#)

Definition

A standard lamp is used as a reference in photometric, radiometric, and colorimetric measurements for which the calibration is traceable to a primary photometric, radiometric, or colorimetric standard. According to the basic and general

concepts and the associated terms of the International Vocabulary of Metrology, the measurement standard is routinely used to calibrate or verify measuring instruments or measuring systems. In color applications, the standard lamp is the measurement standard that transfers the quantities from the reference measurement standard in a metrological organization [1, 2].

Overview

Colorimeters and spectrometers need to be calibrated against a standard lamp with known standardized data. The International Commission on Illumination (CIE) standard illuminant A is intended to calibrate typical, domestic, tungsten-filament lamps. Its relative spectral power distribution is that of a Planckian radiator at a temperature of approximately 2856 K according to the International Temperature Scale of 1990 (ITS-90) with $c_2 = 14,388 \mu\text{m}$. Also, the definition of CIE source A is used to realize the CIE illuminant A. CIE standard illuminant A can be realized by CIE source A, defined as a gas-filled, tungsten-filament lamp operating at a correlated color temperature of 2856 K. The standard lamp can be a CIE source A to transfer the quantities from the reference measurement standard [3].

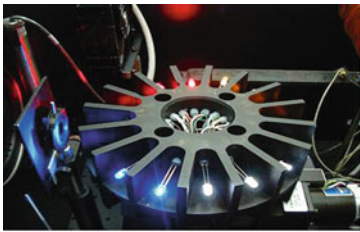
To establish the traceability of a measurement system, a working standard lamp with the measurement data is assigned by the substitution procedure to compare the primary standard. A working standard lamp is usually calibrated with reference to a secondary standard lamp that is operated under similar conditions and has a certain certification procedure to verify the performance of quantities with a degree of uncertainty. In general, the working standard lamp is applied for the transfer of photometric and radiometric quantities, such as luminance, irradiance, luminous intensity, luminous flux, CIE chromaticity coordinates, etc. The stable and repeatable transfer standards with a low degree of uncertainty, which have similar characteristics to those of the test lamp, are desired for the measurement system

to transfer photometric and colorimetric units from a national metrology institution (NMI).

Some NMIs are now in charge of developing the light-emitting diode (LED) transfer standard to provide the LED industry with a measurement system that is traceable to SI units, as shown in Fig. 1. Also, the transfer standard carries the photometric and radiometric quantities, which are the alternative standards to the traditional gas-filled lamp.

There are three typical LED standard lamps that have been developed in the NMIs. The traditional LED package is the 5-mm diameter package, which has been used since the early 1960s. The LED die is mounted in the reflector cavity, which bonds two wires from the LED die and connects to the frame. There are many colors of LED packaged in transparent epoxy. Some white ones are mixed the phosphors in transparent epoxy lens. The typical LED package – a 5-mm lamp LED transfer standard – has been developed by the seasoning system and the correction model of temperature and photometric quantities has been applied [4]. Although LEDs have many advantages over conventional sources in becoming a transfer standard, they generate a large amount of heat. The heat not only influences the optical characteristics of LEDs, but also alters the stability and repeatability. Therefore, it is important to design a thermal control system to overcome the temperature problem. The high-powered LED should have a suitable temperature control module to reduce the high temperature on the LED die. Therefore, some NMIs have developed a high-powered LED with a thermal control module for the LED standard lamp that has two light distributions for different applications [5]. We can use the goniophotometer measurement method to provide stable photometry and colorimetry quantities for producing the high-powered LED standard lamp [6].

In addition, some NMIs have developed controllable light distribution and have a lower angular correlated color temperature distribution (ACCTD) LED transfer standard for the various applications of LED measurement systems [7]. Also, NMIs provide the temporal and luminous



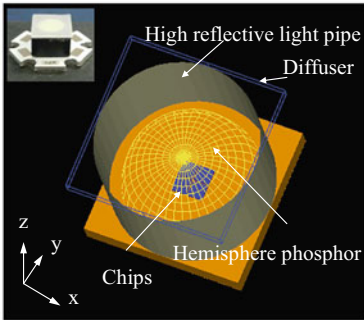
Typical packaging of an LED standard lamp



An LED standard lamp



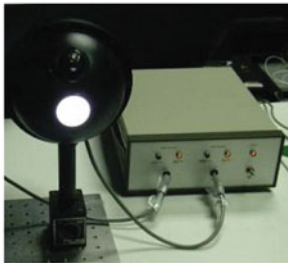
High-powered LED standard lamp



Heat-sink free high-powered LED lamp



High-powered LED bulb standard lamp



Temporal and luminous standard lamp

Standard Lamp, Fig. 1 Various light-emitting diode (LED) transfer standards developed by national metrology institutions

standard lamp for the display measurement instrument [8].

Application (For Example, How to Use the Standard Lamp, Spectral Radiometer)

We can use the spectroradiometer to measure the spectral power distribution of the luminaire. The typical sensor of the spectroradiometer for the visible wavelength is silicon-based. The spectral response is not the same for the measurement range. Moreover, the spectral efficiency of the grating is not the same either. Therefore, we should use the standard lamp to calibrate the spectroradiometer using the replacement method. The standard lamp carries the spectral irradiance standard for the measurement of spectral range calibrated by the metrological institute. We use the spectroradiometer to measure the standard lamp, then correct the

spectral response of the spectroradiometer. Furthermore, using the gas lamp (Xe, Ar, Ne, and Kr), which has distinct spectral lines, the spectral reference calibrates the spectroradiometer for wavelength correction. Therefore, we can use the calibrated spectroradiometer to measure the device under test (DUT).

Cross-References

- [Spectral Power Distribution](#)
- [Spectroradiometer](#)

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. The Joint Committee for Guides in Metrology.: International Commission on Illumination: International Vocabulary of Metrology – Basic and General Concepts

- and Associated Terms. JCGM 200:2012, 3rd edn. BIPM, Sevres Cedex, France (2012)
3. János, S. (ed.): *Colorimetry: Understanding the CIE System*. Wiley, Hoboken (2007)
 4. Park, S.C., Kim, Y.W., Lee, D.H., Park, S.N.: Preparation of a standard light-emitting diode (LED) for photometric measurements by functional seasoning. *Metrologia*. **43**, 299–305 (2006)
 5. Gerloff, T., Lindemann, M., Shirokov, S., Taddeo, M., Pendsa, S., Sperling, A.: Development of a new high-power LED transfer standard. In: *Proceedings of CIE 2012 Lighting Quality & Energy Efficiency*, Hangzhou, China, pp. 125–127 (2012)
 6. Zama, T.: Developing a new photometric standard for light emitting diode (LED) – ensuring reliable photometric evaluation of LED. *AIST Today*. **40**, 17 (2011)
 7. Chen, C.H., Chang, Y.Y., Ting, Z.Y., Sun, C.C.: Design of a standard LED light source with extremely high optical and color stability. In: *Frontiers in Optics*, Orlando (2013)
 8. Liu, W.C., Wu, G.N., Kuo, C.J., Chen, Y.L.: A light standard for FPD dynamic parameter. In: *Secretariat of Asia Display 2011 Conference*. Kanshan, China (2011)

Standard Measurement Geometries

Bor-Jiunn Wen¹ and Chao-Hua Wen²

¹Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Standard illuminating and viewing conditions](#)

Definition

Reflectance and transmittance measurements depend in part on the geometry of illumination and viewing. Accordingly, the CIE has specified illuminating and viewing conditions, namely, 45°:0°a, 45°:0°c, 45°:0°x, 0°:45°, d:0°, 0°:d, de:8°, and di:8°. With the 45°:0° geometry, the

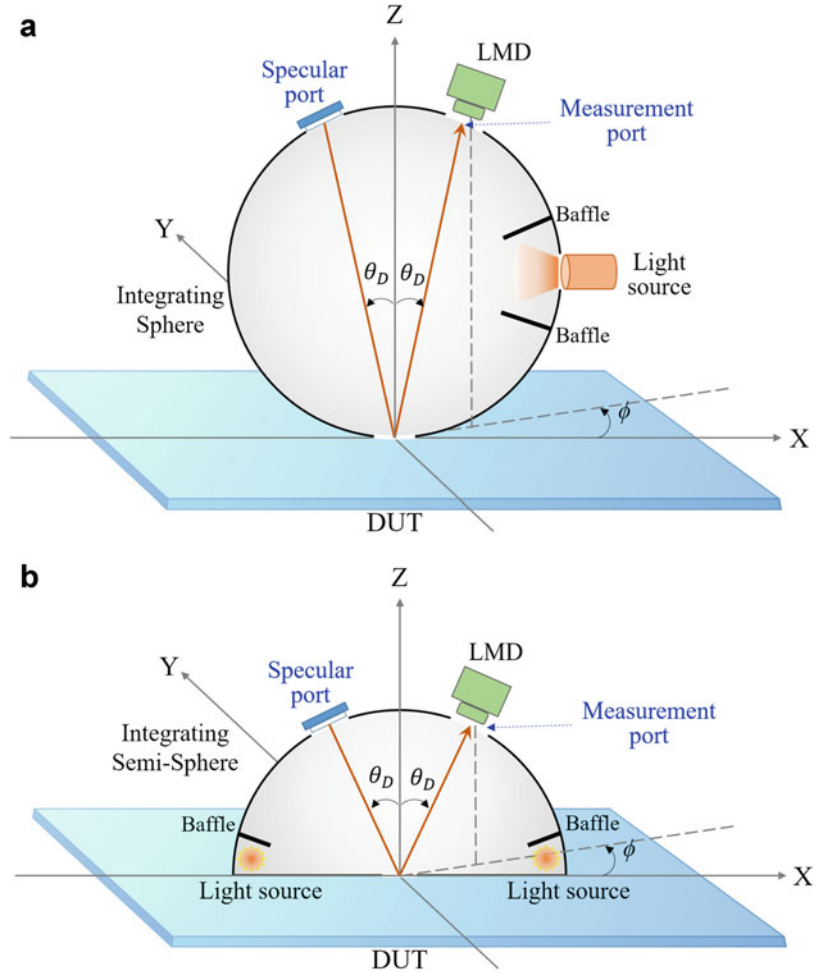
sample is illuminated at an angle of 45° from the normal to the sample surface and viewed perpendicularly to the sample surface or within 10° to the normal. There are three standard configurations of the geometry: 45°:0°a, 45°:0°c, and 45°:0°x. The 45°:0°a has a 10° wide cone of light incident on the sample at 45° to the sample normal. The “a” in 45°:0°a indicated annular meaning that the 10° cone continues uninterrupted all the way around the sample. The 45°:0°c indicates circumferential illumination, which is similar to annular except that it is distributed in discrete locations around the ring. The 45°:0°x geometry uses a single beam at a single azimuth angle. The 0°:45° geometry is the reverse of 45°:0°. The 0°:d (normal/diffuse) geometry is one in which the sample is illuminated by a beam whose axis is at an angle not exceeding 10° from the normal to the sample surface. The reflected flux is collected by an integrating sphere, and viewing is accomplished by a receiver being directed toward an inner wall of the sphere. In all geometries, the viewing cones should not exceed 5°. The d:0° geometry is the reverse of 0°:d. The d:8 hemispherical geometries present the sample with diffuse illumination and approximately normal detection. Diffuse illumination is accomplished by the use of an integrating sphere. Light introduced into the sphere is sent through a series of reflections off the diffuse white sphere wall, quickly resulting in the diffuse illumination incident on the sample. Two standard configurations of diffuse geometry are de:8° and di:8°. The former excludes the specular component and the latter includes it.

Example of Measurement Geometry

Diffuse and directional illumination geometries are used for determining the color performance of the device under test (DUT). This entry introduces the examples of measurement geometries for flat panel displays. In practice, configurations for implementing these illumination geometries are represented as follows [1–6].

Standard Measurement Geometries, Fig. 1 (a)

Type I and (b) type II



Diffuse Illumination

Hemispherical Illumination

Use a light source with an integrating sphere, as a uniform diffuse light source, to illuminate DUT. Place the light-measuring device (LMD) at the measurement port with an inclination angle of θ_D off the normal to the center of the DUT. (See Fig. 1. When θ_D is 8° , the measurement geometry is d:8. The 8:d geometry is the reverse of d:8. The d:8 and 8:d geometries are specified by the CIE as standard viewing/illuminating geometries.) The measurement area covers no less than 500 pixels, and measurement field angle is 2° or less. For the specular reflection excluded measurement, open the cover of the specular port during the test. For the specular reflection

included measurement, close the cover of the specular port during the test.

Directional Illumination

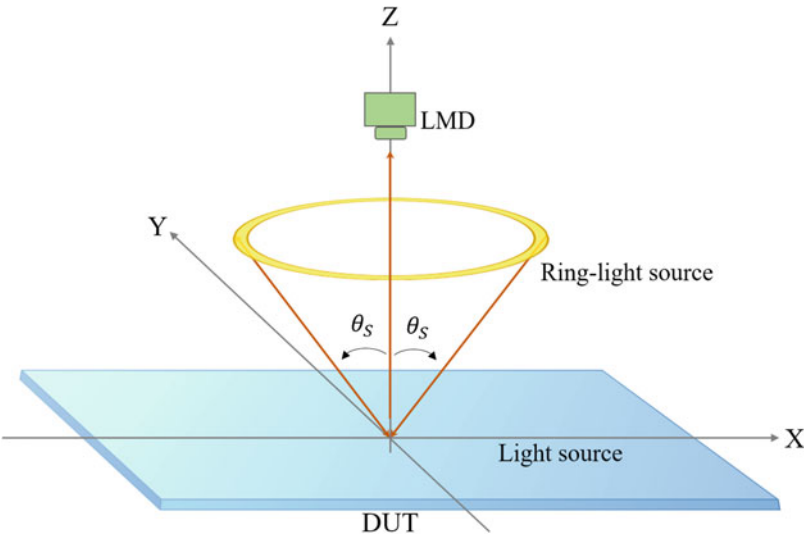
Ring Light Illumination

Use a ring light source to provide uniform directional illumination on the DUT over all azimuth angles at an inclination angle of θ_S (recommend 45°). Place the LMD normal to the center of the DUT (see Fig. 2). The measurement area covers no less than 500 pixels, and the measurement field angle is 2° or less.

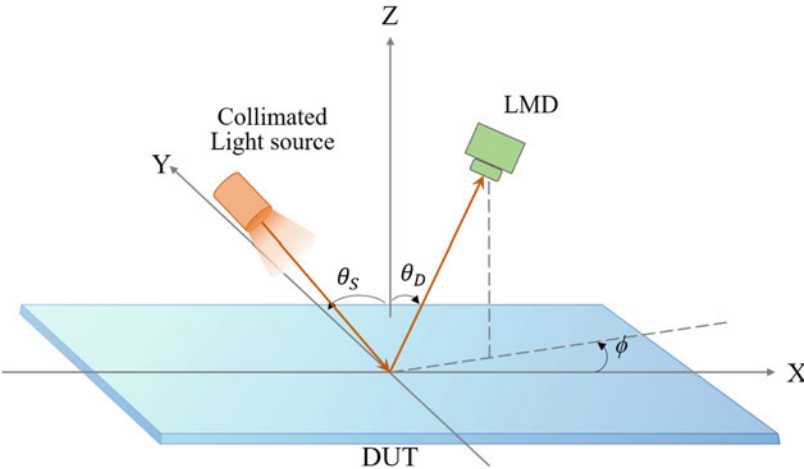
Collimated Light Illumination

Use a collimated light source to provide directional illumination on the DUT at an inclination

Standard Measurement Geometries, Fig. 2 Ring light illumination measurement geometry



Standard Measurement Geometries, Fig. 3 Collimated illumination measurement geometry



angle of θ_s . Place the LMD at a viewing angle of θ_D . (See Fig. 3. When θ_s is 0° and θ_D is 45° , the measurement geometry is 0:45. The 45:0 geometry is the reverse of 0:45. The 0:45 and 45:0 geometries are also specified by the as CIE standard viewing/illuminating geometries.) The measurement area covers no less than 500 pixels, and the measurement field angle is 2° or less.

Application 1

Users usually view liquid crystal displays (LCDs) in a bright environment, such as office, living

room, etc. To evaluate the contrast of display devices under ambient light is more practical than that under a dark environment. The results of ambient contrast measurements are affected by different illuminance, view angles, etc. Therefore, the measurement geometry of the ambient illuminance is important for the measurement of ambient contrast for LCDs [7].

For the diffuse illumination, a uniform diffuse-ambient light is provided by a light source with an integrating sphere to illuminate the screen of the display (see Fig. 1). The integrating sphere has plural ports. DUT is placed against the sampling port of the integrating sphere. The light

measurement display LMD is arranged to view the screen through the measurement port. Only one measurement port is in use during the measurement. Measurement ports not in use have covers on them. The inner surfaces of covers are coated with standard white that should have the same reflection properties with the inner surface of the sphere. In addition, the directional illumination can also be provided to illuminate the screen of the display for the measurement of ambient contrast of LCDs.

After setting up the measurement geometry, the procedures of the measurement of ambient contrast of LCDs by using the diffuse illumination are listed as follows.

For the measurement with the specular port closed, warm up the DUT for at least 40 min firstly. Secondly, measure the luminance of full-screen white $(L_W)_{dark}$ and full-screen black $(L_K)_{dark}$ under darkroom conditions. Thirdly, set up the integrating sphere, and turn on its lamp (warm up for at least 30 min). Fourthly, measure the illuminance on the surface of DUT. The illuminance is adjusted to the required levels (200 and 500 lx are typical for a bright living room and an office environment) by controlling the intensity of the light source. It is important to keep the correlated color temperature (CCT) of the light source stable during the illuminance adjustment. Finally, measure the luminance of white (L_W) and black (L_K) of full-screen patterns. The diffuse illumination should be kept stable during measuring processes.

For the measurement with the specular port open, warm up the DUT for at least 40 min firstly. Secondly, measure the luminance of full-screen white $(L_W)_{dark}$ and full-screen black $(L_K)_{dark}$ under darkroom conditions. Thirdly, set up the integrating sphere, turn on its lamp (warm up for at least 30 min), and take out the cover of the specular port. Fourthly, measure the illuminance on the surface of DUT. The illuminance is adjusted to the required levels (200 and 500 lx are typical for a bright living room and an office environment) by controlling the intensity of the light source. It is important to keep the CCT of the light source stable during the illuminance

adjustment. Finally, measure the luminance of full-screen white (L_W) and full-screen black (L_K) . The illuminance should be kept stable during the measurement process.

After the measurement of ambient contrast of LCDs, the ambient contrast can be directly calculated by using Eq. (1), where C_A is the symbol of ambient contrast and L_W and L_K are the luminance of full-screen white and full-screen black, respectively.

$$C_A = \frac{L_W}{L_K} \quad (1)$$

The ambient contrast can also be calculated by using the reflectance. The reflectance of full-screen white ρ_W and full-screen black ρ_K is derived by Eq. (2), and the ambient contrast C_A is calculated by Eq. (3). It is recommended that the reflectance is measured at the illuminance E_{01} not smaller than 1000 lx. In Eq. (3), E_0 is the desired illuminance (200 and 500 lx are typical for a bright living room and an office environment).

$$\begin{cases} \rho_W = [L_W - (L_W)_{dark}] \times \frac{\pi}{E_{01}} \\ \rho_K = [L_K - (L_K)_{dark}] \times \frac{\pi}{E_{01}} \end{cases} \quad (2)$$

$$C_A = \frac{(L_W)_{dark} + \frac{\rho_W}{\pi} E_0}{(L_K)_{dark} + \frac{\rho_K}{\pi} E_0} \quad (3)$$

Application 2

For solid-state lighting luminaries (SSL), the standards are described in the framework of the Energy Star program. One of the standards therein is the IES LM-79 on testing and measuring methods. Hereunder, the measurement of the luminous flux and of the luminous colors using integrating sphere photometers and their related requirements is described. For measurement geometries, the LM-79 recommended by integrating sphere has two kinds: 2π sphere and 4π sphere. In the 2π version, light sources are

operated outside the sphere. The luminous flux emitted from the light source enters the sphere through a measuring aperture. The application range is limited to spotlights of the maximum 2π (hemisphere). In the 4π version, the light sources are operated at the center of the integrating sphere with which all directional radiation can be measured. In the LM-79, it is recommended to measure 2π and 4π sources in the inner part of the sphere. The use of the 2π sphere geometry is only recommended in exceptional cases, i.e., when required due to the dimensions and form of the light source. Therefore, further key concerns of the LM-79 with the integrating sphere are:

- Reflectivity of the sphere coating
- Field of view function of the detectors
- Baffle implementation
- Temperature measurement ability

It is very important to check the key concerns for SSL measurements before using the LM-79 standard.

Cross-References

- [Illuminance Meter](#)
- [Luminance Meter](#)

References

1. Kuehni, R.G.: Color: An Introduction to Practice and Principles, 3rd edn. Wiley, Hoboken (2012)
2. Fairchild, M.D.: Color Appearance Models, 3rd edn. Wiley, Chichester (2013)
3. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd edn. Wiley-Interscience, Hoboken (2000)
4. Hunt, R.G.W.: The Reproduction of Colour, 6th edn. Wiley, Chichester (2007)
5. SID/ICDM: Information Display Measurements Standard, v1.03 (2012)
6. SEMI D68-0719: Test methods for optical properties of electronic paper displays (2019)
7. SEMI D56-0519: Test method for measurement for ambient contrast of liquid crystal displays (2019)

Stanziola, Ralph A

Michael H. Brill

Datacolor, Lawrenceville, NJ, USA



Ralph A. Stanziola (1931–2007), a consultant and teacher of color technology, was born in Philadelphia, received his B.S. degree in Chemistry from the Philadelphia College of Textiles and Science (now Philadelphia University), and resided most of his life in New Jersey. In his early years, at the Research and Technical Service for the Dyes Department of the American Cyanamid Company, Ralph learned color technology from such pioneers as Orrin W. Pineo and Edward I. Stearns. He next served for 9 years as Technical Representative and General Sales Manager for the Davidson & Hemmendinger Company, and later became a sales manager for the Kollmorgen Corporation Color Systems division, which had acquired Davidson and Hemmendinger.

In 1970, Ralph cofounded Applied Color Systems, Inc. where he eventually became the Executive Vice President and Technical Director. In May 1985, Ralph founded Industrial Color Technology, a consulting company under whose auspices he solved many industrial problems involving color control. For the rest of his life, Ralph remained a consultant for ACS (later Datacolor), instructed their Color Technology

seminars, and helped to create the first video training series on color technology.

Ralph held five US patents [1–5] and authored a large number of technical papers and presentations. He developed the Color Curve System for color communication, and helped the Glad Products Company to develop the now famous “Glad Difference,” *Yellow and Blue Make Green*® seal. Less familiar will be his development (with Bob Swain of Chroma Corp.) of a method of using a colored sample to test the wear of a metal piece. The test consists of extruding a molded plastic piece made of component materials of two different colors, and monitoring the color of the mixture. Earlier, Ralph developed the first colorant-dispenser system driven by computer color matching (1979) and patented a Maxwell-disk-based color simulator (1980). He also codeveloped an asymmetrical color tolerancing system using artificial intelligence algorithms (1991), a photonic visual color simulator using LEDs (late 1990s), and a new version of the Color Rule for testing observers and light booths for metamerism (2006).

Ralph joined the Inter Society Color Council (ISCC) in 1962 and was an active member for the rest of his life. He cochaired several ISCC meetings: the Annual Meeting in Princeton in 1992 that celebrated the 25th anniversary of the AIC; the 1994 joint meeting with the Detroit Color Council (DCC); and the 2003 Williamsburg Conference on Industrial Color Problems held at Philadelphia University. He was instrumental in setting up the Education Interest Group, and was the first to chair the Industrial & Applied Color Interest Group. In recognition of these contributions, Ralph received the 2004 Nickerson Service Award, and in 2005 became an Honorary Member of the ISCC. He also served as *Color Research and Application's* Special Editor for Industrial Applications for 10 years.

In 1995, the Technical Association of the Pulp and Paper Industry presented Ralph with their Finest Faculty Award. As a member of the Federation of Societies for Coatings Technology, he received the Armin J. Bruning Award for his outstanding contribution to the science of color in the field of coatings technology. He was also a member of the American Association of Textile

Chemists and Colorists and the Detroit Colour Council.

Ralph's numerous lectures included computer color matching seminars at the Rensselaer Color Measurement Laboratory (Rensselaer Polytechnic Institute) and at the Munsell Color Science Laboratory (Rochester Institute of Technology). The many personal anecdotes he told in his color courses showed a wealth of experience not available in any book. Some urged him to write a book, but he preferred the humbler route of interacting directly with the world and with other people.

References

1. Worn, P.R., Stanziola, R.A.: Hall D.R. Reflected color simulator. US Patent 4,310,314 (1982)
2. Stanziola, R.: Objective color notation system. US Patent 5,012,431 (1991)
3. Swain, R.D., Stanziola, R.A.: Method of determining wear. US Patent 6,306,319 (2001)
4. Swain, R.D., Stanziola, R.A.: Method of determining wear. US Patent 6,797,204 (2004)
5. Riddle, G.H.N., Reitmeier, G., Steinmetz, C., Stanziola, R.A., Burstyn, H.C.: Color display device. US Patent 6,985,163 (2006)

Stereo and Anaglyph Images

Krzysztof Templin

Computer Graphics, Max Planck Institute for Informatics, Saarbrücken, Germany

Synonyms

[Stereo pairs](#); [Stereograms](#)

Definition

Stereo images are pairs of conventional images depicting a scene as seen from two different viewpoints, which correspond to the eye positions of an observer in the scene. When each image of a stereo pair is presented separately to the corresponding eye, the impression of three-dimensionality known as stereopsis arises. Stereo

images find its use in arts, entertainment, and scientific visualization.

An anaglyph image is a method of presentation of a stereo image, in which the images composing the stereo pair are colored with different hues and superimposed. The resulting image is viewed using glasses with matching color filters, so that each eye sees only one of the images and stereopsis can appear. Usually, red and cyan filters are used.

Stereo and Anaglyph Images

Stereo Images

Since human eyes have different vantage points from which they view the world, a point in space may be projected to different positions in the two eyes. The difference between the points of projection is called “binocular disparity” and it is a basis to the impression of depth called “stereopsis” [1]. Stereopsis can be artificially triggered when two images, in which corresponding points have been displaced according to their depth, are presented dichoptically, i.e., one to the left eye and the other to the right eye. The images (called half images) are fused into a percept of a single, three-dimensional image. An important tool for studying stereopsis is a random dot stereogram, where the depth impression is evoked solely by the binocular disparity in isolation from any other sources of depth information [2].

The simplest method of stereo image presentation is free fusion. The half images are shown side by side, and the observer is expected to simply converge or diverge their eyes, until the images overlap. The invention of the first device for watching stereograms is attributed to Wheatstone and dates back to 1838 [1, 3]. The modern presentation methods range from various setups making use of special glasses (color, polarizing or interference filters, time multiplexing) to glasses-free, autostereoscopic displays. Each method has its advantages and limitations, important factors being usability, viewing comfort, and faithful color reproduction. A serious flaw of a stereo system is cross talk, i.e., leakage of the signal intended for one eye to the other eye. Cross talk hinders fusion and in extreme cases can spoil the 3D effect completely.

An important issue in stereoscopy is viewing comfort [4]. Since stereo images are presented in a single plane, they require decoupling the focusing and vergence mechanisms of the observer’s eyes. This restricts the range of depths possible to reproduce to a certain volume around the display, because excessive mismatch between the focusing and vergence points causes significant eye-strain in the observer [5]. Since more distant objects require less refocusing, the comfort zone can be greatly extended by increasing the distance to the display or by using lenses which move back the focusing point. Other sources of discomfort include, but are not limited to, various types of inconsistencies between the half images, e.g., in vertical position, focus, or color balance.

When capturing a stereo image, the distance between the viewpoints does not need to equal the human interocular distance (ca. 64 mm). One situation when this is useful is capturing very distant or very close objects. For the former, there would be hardly any stereopsis if natural separation was used, while for the latter the resulting image could become impossible to fuse. Altering the separation between the viewpoints may be also necessary to maintain plausible proportions of the objects or to fit the content within the comfort zone of the display. Sometimes, the background is captured using different viewpoint separation than the foreground and the outcome is composed into one image [6], or the disparities are adjusted throughout the whole range of depths using image processing techniques [7].

Anaglyph Images

Anaglyph images, first described in 1853 [8], are probably the most widespread presentation method of stereo images. The half images are colored with different colors and superimposed using additive color mixing. Each filter of the glasses lets in the light from the corresponding half image while blocking the light from the other one. Various color combinations are used, such as green-magenta or red-blue; however, the most common one is red-cyan, with the left filter transmitting red light and the right filter transmitting blue and green. Apart from the glasses, this method does not require any special equipment,

Stereo and Anaglyph

Images, Fig. 1 A pair of paper anaglyph glasses with red and cyan filters



Stereo and Anaglyph Images, Fig. 2 A grayscale stereo image presented as an anaglyph image; from left to right: red-cyan, green-magenta, yellow-blue

such as a dedicated printer or display, which makes it very inexpensive and convenient (see Fig. 1). However, since the colors of the image need to be modified, this method suffers from poor color reproduction. Moreover, each half image is manipulated in a different way, so the outcome may cause retinal rivalry. Slight blurriness due to chromatic aberration may be visible in one of the eyes, and some high-end glasses include corrective lenses to counteract this effect. Finally, since the filters often do not separate the two half images perfectly, cross talk may appear.

In the most basic form of anaglyph 3D for RGB displays, one filter transmits one of the primaries and the other filter the remaining two primaries. Hence, three types of anaglyph images are possible: red-cyan, green-magenta, and blue-yellow. The final image is formed by selecting one of the RGB channels from one of the half

images and the remaining two channels from the other (see Fig. 2). The yellow filter transmits most of the perceived colors; however, there is a significant luminance imbalance between the yellow and blue filters. The green-magenta pair is well balanced in terms of luminance but gives rise to noticeable color sheen over a wide range of hues [9]. The colors can be traded off for viewing comfort by converting one or both half images to grayscale or by ignoring one of the channels completely [10].

Several algorithms have been proposed to enhance the quality of the anaglyph image given a pair of filters. If the spectral absorption curves of the glasses and the density functions of the display primaries are known, the creation of an anaglyph image as close as possible to the desired stereo pair can be posed as an optimization problem [11]. By performing a few color matching tasks,

anaglyph image generation process can be adjusted to reduce ghosting [12]. Proprietary systems exist, such as ColorCode 3-D or Inficolor 3D, that combine advanced image processing techniques with careful filter pair selection.

Further improvement in the color reproduction can be achieved by using wavelength multiplexing. Each filter transmits all three primary colors but in slightly different wavelength subbands for each eye. The half images have different gamuts, one leaning toward green and the other toward magenta, but the visual system compensates for the difference, and after a short time it is not visible anymore. Even better color fidelity can be obtained with systems based on light polarization. The disadvantage of both systems in comparison to anaglyph 3D is that in addition to the glasses, they require special displays [6, 10].

Cross-References

- [Chromostereopsis](#)
- [Trichromacy](#)

References

1. Howard, I.P., Rogers, B.J.: *Perceiving in Depth*. Oxford University Press, New York (2012)
2. Julesz, B.: *Foundations of Cyclopean Perception*. MIT Press, Cambridge, MA (2006)
3. Wheatstone, C.: Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Phil. Trans. R. Soc.*, vol. **128**, pp. 371–394, London (1838)
4. Lambooi, M., Fortuin, M., Heynderickx, I., IJsselstein, W.: Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol.* **53**(3), 1–12 (2009)
5. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* **8**(3), 33.1–30 (2008)
6. Mendiburu, B.: *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, Boston (2012)
7. Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.* **29**(4) (Proc. ACM SIGGRAPH), 75:1–10 (2010)
8. Rollmann, W.: Zwei neue stereoskopische Methoden. *Ann. Phys.* **166**, 186–187 (1853)
9. Sorensen, S.E.B., Hansen, P.S., Sorensen, N.L.: Method for recording and viewing stereoscopic images in color using multichrome filters. US Patent No. 6,687,003, 2004
10. Hainich, R.R., Bimber, O.: *Displays: Fundamentals and Applications*. A K Peters/CRC Press, Boca Raton (2011)
11. Dubois, E.: A projection method to generate anaglyph stereo images. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 1661–1664 (2001)
12. Sanftmann, H., Weiskopf, D.: Anaglyph stereo without ghosting. *Comput. Graphics Forum* **30**(4) (Proc. EGSR), 1251–1259 (2011)

Stereo Pairs

- [Stereo and Anaglyph Images](#)

Stereograms

- [Stereo and Anaglyph Images](#)

Stevens, Stanley Smith

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



S.S. Stevens was an American experimental psychologist perhaps best known in the world of color science for introducing the psychophysical power law and for collecting data on brightness perception as a function of adaptation [2]. He was even more well known in the general field of experimental psychology and psychophysics for many accomplishments in acoustical psychophysics, his important publications, and founding of the Psycho-Acoustic Laboratory at Harvard University.

Among Stevens' literary contributions to the field are the extensive, and immensely praised, *Handbook of Experimental Psychology* (1951) [3] that provides much information that remains useful more than half a century later and the text *Psychophysics* (1975) [4] that was published posthumously after final editing by his wife. It, too, remains a useful reference on the topic. Among the topics described is the hierarchy of mathematical scales (nominal, ordinal, interval, and ratio) in psychophysics, something else developed by Stevens.

Stevens was an undergraduate student at the University of Utah and then transferred to Stanford where he completed his degree in an undetermined field since the variety of courses he took was so extensive. He then was accepted at the Harvard Medical School, but chose to enroll in Harvard's School of Education to avoid the medical school's \$50 fee and organic chemistry requirement and still have access to Harvard's resources. At Harvard he was transformed by a course on perception taught by E.G. Boring, became Boring's unpaid research assistant, and earned a Ph.D. in Philosophy by the end of his second year there. After some more studies in physiology and physics, he was appointed an instructor in the Psychology Department, where he remained until his death. In 1962, Harvard University agreed to appoint Stevens as Professor of Psychophysics in honor of the field founded by G.T. Fechner.

One of Stevens' most notable papers was entitled "To honor Fechner and repeal his

law," published in *Science* in 1961 [1]. In it, Stevens summarized psychophysical scaling data for a variety of perceptual stimuli (e.g., brightness of light, loudness of sound, hotness of heat, pain of electric shock, etc.) and illustrated that they all could not possibly follow the logarithmic relationship between stimulus and perception predicted by Fechner's Law. Instead, all of his data could be well described by power functions with various exponents, which depend on the perception being scaled. This new relationship is known as the psychophysical Power Law, or the Stevens' Power Law. One technique used to establish the Power Law that Stevens developed is known as cross-modal matching. In these experiments, observers match perceived magnitude of one stimulus (e.g., brightness of light) with another (e.g., loudness of sound).

Stevens' work on brightness is well known in color science and led to the description of a color appearance phenomenon as the Stevens' Effect. This work, published in 1963 by Stevens (no relation) and Stevens, examined the effects of light and dark adaptation on the perception of brightness. Scales of brightness were obtained using magnitude estimations (a Stevens' hallmark) and showed that the contrast of brightness increased with increasing adapting luminance (exponent in the power increases). So, as Stevens showed us, when the level of lighting increases, dark colors look darker and light colors lighter.

References

1. Stevens, S.S.: To honor Fechner and repeal his law. *Science*. **133**, 80–86 (1961)
2. Stevens, J.C., Stevens, S.S.: Brightness functions: effects of adaptation. *J. Opt. Soc. Am.* **53**, 375–385 (1963)
3. Stevens, S.S. (ed.): *Handbook of Experimental Psychology*. Wiley, New York (1951)
4. Stevens, S.S.: *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. Wiley, New York (1975)

Stiles, Walter Stanley

Robert W. G. Hunt

Department of Colour Science, University of
Leeds, Leeds, UK



Dr. W.S. Stiles, OBE, FRS, trained in physics at University College, London, and in mathematics at St. John's College, Cambridge. He spent almost all his working life at the National Physical Laboratory of Great Britain in Teddington.

He was created OBE (Officer of the British Empire) in 1946 for his wartime work on visibility, visual search, and the defensive use of dazzle. He was elected to the (British) Royal Society in 1957 and was awarded the Tillyer Medal of the Optical Society of America in 1965. He served as General Secretary of the Commission Internationale de l'Éclairage from 1928 to 1931, was Chairman of the Colour Group of Great Britain from 1949 to 1951, and was President of the (British) Illuminating Engineering Society from 1960 to 1961.

His earlier work with B. H. Crawford introduced the concept of veiling glare, a subject of

particular importance in street lighting, and other applications of illuminating engineering.

His name is perhaps best known in the Stiles-Crawford effect, the reduction in sensitivity of the retina as the angle of incidence of the light becomes increasingly different from normal.

In the late 1950s, he spent much time and effort in constructing a visual colorimeter for redetermining the color-matching functions of the normal observer. This work, involving measurements on over 50 observers, provided the major basis for what became the CIE 1964 supplementary standard colorimetric observer for field sizes of 10° ; the study also included experiments with a 2° field which confirmed the validity of the CIE 1931 standard colorimetric observer apart from the well-known deficiency in sensitivity at the extreme short-wave end of the spectrum. This work also addressed the phenomenon of rod intrusion: the fact that in fields of 10° size, some input from the rods can be added to that of the cones.

Much of his later work was devoted to the study of increment thresholds: the magnitude of the just-noticeable amount of a stimulus of one wavelength when superimposed on a uniform field of another wavelength. From these studies he identified a series of basic visual mechanisms which he identified as π mechanisms, the significance of which have been a matter for considerable discussion.

Toward the end of his life, he collaborated with Gunter Wyszecki in the production in 1967 of *Color Science: Concepts and Methods, Quantitative Data and Formulas*, published by Wiley, with a second edition in 1982 [1]. This work still provides an enormous amount of useful data and information on color science.

He was dedicated to traditional psychophysical experimental methods and regarded it as very necessary to avoid arguments founded on introspective descriptions of sensations, which he regarded as notoriously difficult to interpret correctly. He therefore had no contact with the area of magnitude estimation pioneered by S.S. Stevens and developed subsequently by other workers in color science.

As a lecturer he was clear and authoritative with a commanding demeanor. Although a man of great intellect and prodigious experience in color science, he was nevertheless always willing to help others in the field who approached him for help. His interests included mathematics, reading, and painting.

References

1. Wyszecki, G., Stiles, W.: *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. Wiley, New York (2000)

Further Reading

- Stiles, W.S.: The directional sensitivity of the retina and the spectral sensitivities of the rods and cones. *Proc. Roy. Soc. Lond.* **B127**, 64 (1939)
- Stiles, W.S., Burch, J.M.: Interim report to the Commission Internationale de l'Eclairage, Zurich (1955), on the National Physical Laboratory's investigation of colour matching. *Optica. Acta.* **2**, 168 (1955)
- Stiles, W.S., Crawford, B.H.: The luminous efficiency of rays entering the eye pupil at different points. *Proc. R. Soc. Lond B.* **112**, 428–450 (1933)

Street Lighting

► Road Lighting

Stroop Effect

Colin M. MacLeod
Department of Psychology, University of
Waterloo, Waterloo, ON, Canada

Synonyms

[Color-word interference](#); [Stroop interference](#)

Definition

The Stroop effect is one of the best known phenomena in all of cognitive science and indeed in

psychology more broadly. It is also one of the most long standing, having been reported by John Ridley Stroop in the published version of his dissertation in 1935 [1]. In its basic form, the task is to name the color in which a word is printed, ignoring the word itself. When the word is a color word printed in a mismatched ink color, this is very difficult to do and results in slow, error-prone responding. To illustrate, consider **GREEN**: To say “red” to the ink color is difficult relative to a variety of comparison or control conditions such as naming the color of **XXXXX** or the word **TABLE** or even the word **RED**. The performance cost in the mismatch condition – usually referred to as the incongruent condition – relative to the controls is called the Stroop effect or Stroop interference. Figure 1 provides an illustration of the phenomenon.

Background

Since the very beginning of experimental psychology, it has been clear that words are faster to read than objects or their properties are to name. In his dissertation in 1886, Cattell [2] even went so far as to suggest that word reading is automatic due to extensive practice, introducing the concept of automaticity to cognitive science. Automatic processes can be thought of as unintentional, uncontrolled, unconscious, and fast [3]. Under the automaticity account, people cannot comply with an instruction not to read because reading cannot be “turned off”; hence it is guaranteed that incompatible words will cause interference when attempting to name their print colors.

From early on, the other prevalent explanation of Stroop interference was the relative speed of processing account, which in its simplest form argued that faster processes can affect slower processes but not vice versa [4]. Thus, because words are read faster than colors can be named, interference results when the task is to name the colors and ignore the words. This also fits nicely with Stroop's other finding – that there was no “reverse Stroop” interference when the task was to read the words and ignore the colors: Reading performance for incongruently colored words was

1	2	3	4
red	red	xxx	yellow
yellow	yellow	xxxxxxx	blue
green	green	xxxxxx	red
red	red	xxx	yellow
blue	blue	xxxx	red
green	green	xxxxxx	blue
yellow	yellow	xxxxxxx	green
blue	blue	xxxx	green
green	green	xxxxxx	red
red	red	xxx	blue
blue	blue	xxxx	yellow
yellow	yellow	xxxxxxx	green

Stroop Effect, Fig. 1 An illustration of the Stroop effect. In columns 1 and 2, the task is to read each word in the column aloud, ignoring its print color, and to do so as quickly as possible. This represents Stroop’s (1935) first experiment, where he found little difference in reading time between the experimental condition (column 2) and the control condition (column 1). In columns 3 and 4, the task is to name the print color of each word in the column aloud, ignoring the word itself, again doing so as quickly as possible. This represents Stroop’s (1935) second experiment, where he found dramatic interference in that color naming time in the experimental condition (column 4) was much slower and more error prone than was the case in the control condition (column 3). Note that to make the conditions comparable, the same responses are required for every item in every column

equivalent to that for words printed in standard black ink. Stroop’s results are highly replicable, as MacLeod [5] showed over a half century later.

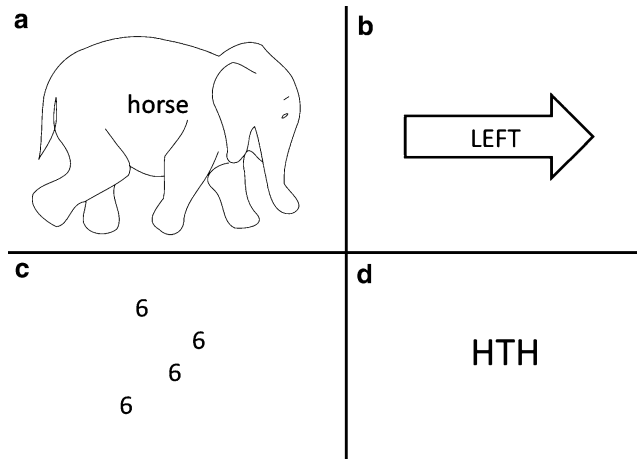
Since Stroop’s landmark study, many hundreds of studies have sought to understand this superficially simple phenomenon, and many more have used his method to explore key aspects of attention, learning, memory, reading, language, and other cognitive skills [5]. More recently, the Stroop task has also been extended to investigate neural mechanisms [6] and clinical disorders [7], among other issues. Interest in Stroop’s method shows no signs of abating; indeed, it is one of the rare phenomena/tasks where interest seems to be growing rather than diminishing with the passage of time.

The Gradient of Interference

An intriguing feature of the Stroop literature, though, is that there was virtually no follow-up to his work for about 30 years; it is only in the 1960s that research on this phenomenon resumed and then with a vengeance. The simplest explanation of this is that the advent of computer-controlled experiments, and especially the resulting ability to time individual trial stimuli, opened up a rich new realm of investigation for which the Stroop task was ideally suited. But the study that actually relaunched research on color-word interference was reported by Klein in 1964 [8] using multiple-stimulus cards in much the same format as Stroop had used. Klein sought to understand what aspects of the words interfered with naming colors. To do so, he incorporated several new conditions. In addition to the baseline “colors-alone” card, there were six interference cards. As always, the standard incongruent condition – using words incongruent with the print colors – showed large interference. When color words that were not the names of the print colors were substituted, interference was cut in half, implicating a large role for response set in interference. Interference then declined more slowly across color-related words (e.g., *lemon*, *sky*); common, unassociated words (e.g., *put*, *heart*); and rare words (e.g., *sol*, *abjure*) and fell to its smallest – but still reliable – level for unpronounceable nonsense syllables (e.g., *hjh*, *gsxrq*). Clearly, despite the instruction to ignore the word, subjects cannot do so, and the greater the relevance of the word, the greater is the resulting interference with color naming. MacLeod [5] reviews the many studies that have taken up where Klein left off.

Variants of the Classic Stroop Task

In essence, Stroop’s paradigm provides a template for studying interference, and investigators have often mined that template to create Stroop-like tasks suited to their particular research purposes. Figure 2 illustrates some of the many alternate



Stroop Effect, Fig. 2 Illustrations of some interference tasks derived from the Stroop “template.” (a) The picture-word task, where naming the picture while trying to ignore the word results in interference (e.g., compared to a picture containing a row of Xs); (b) the “directional Stroop” task, where stating the direction in which the *arrow* is pointing suffers interference from the incompatible word (e.g., compared to no word or a row of Xs); (c) the “counting Stroop”

task, where counting the number of digits suffers interference from the identity of the digits (e.g., compared to using letters in place of the digits); and (d) the flanker task, where identifying the center letter suffers interference from peripheral (flanker) letters that sometimes also appear as central targets (e.g., compared to flankers that are never central targets)

versions in the literature. The best known is the picture-word interference task [9; A in Fig. 2], in which a conflicting word is embedded in a picture. As with the classic Stroop task, interference is largely unidirectional: Naming the picture shows interference from the word, but reading the word is hardly influenced by the picture. Other common variants include the directional version [10; B in Fig. 2], in which again reading the word is quite unhampered by the mismatched arrow, but there is substantial interference from the embedded word when identifying the direction of the arrow, and the digit version [11; C in Fig. 2], in which counting the number of digits is impaired when the digits themselves are incompatible with their numerosity. Further afield, but clearly related, is the flanker task [12, D in Fig. 2], wherein identification of the central letter is impaired by flankers that sometimes also serve as central targets (relative to flankers that are never targets). The list goes on, and MacLeod [5] reviews some of the more common illustrations, although more have appeared in the ensuing quarter century since that review.

Evidence Against the Speed of Processing Account

Stroop himself interpreted the interference phenomenon that now bears his name as follows: “it seems reasonable to conclude that the difference in speed in reading names of colors and in naming colors may be satisfactorily accounted for by the difference in training in the two activities.” Indeed, the third experiment in his paper actually manipulated training and provided support for this explanation, and more recent studies investigating training have confirmed the importance of experience. Notably, MacLeod and Dunbar [13] trained subjects over multiple days to apply color names to initially unfamiliar white shapes. They then were able to track the growth of interference in naming the print colors of incongruent shapes (e.g., the shape called “red” printed in blue) with extended practice on naming just the shapes in white. As shape naming became more practiced (more automatic), shapes produced more interference with naming colors, and print colors produced less interference with naming shapes.

One striking element to the MacLeod and Dunbar pattern was that there was a point in shape training where shape interference in color naming coexisted equally with color interference in shape naming. Such a result is entirely incompatible with a speed of processing account in that the shapes, although still slower to name than colors, nevertheless interfered with color naming. Two converging results were reported around the same time. First, Glaser and Glaser [14] separated the word and color, allowing one dimension to be presented at varying lags before the other. Most critically, even when the color preceded the word by a time sufficient to allow processing of the color before processing of the word, there was still no “reverse Stroop” interference: The color never interfered with the word. Second, Dunbar and MacLeod [15] transformed the words by rotating them in various ways, such as upside down and backwards. This dramatically slowed word reading time, such that it became slower than color naming time. Yet when transformed words were presented in incongruent colors, they still caused as much interference as did upright words. If relative speed of processing were all that mattered for interference, then interference should have been apparent in the Glaser and Glaser studies when color was sped up but not in the MacLeod and Dunbar studies when words were slowed down: In fact, the converse was true. Although compellingly intuitive and widely accepted [4], then, the speed of processing account does not provide a sufficient explanation of Stroop interference.

Alternative Explanations of Stroop Interference

So how *are* we to explain this apparently powerful and simple effect? In the past 25 years, three major explanations have emerged. The first of these was Cohen, Dunbar, and McClelland’s parallel distributed processing, or connectionist, model, proposed in 1991 [16]. At its core, their theory is a strength theory, designed as it was to capture the training data reported by MacLeod and Dunbar [13]. Processing pathways gain strength with

practice, and relative strength determines likelihood and degree of interference. Thus, given our extensive experience with reading, color-word pathways ordinarily are much more strongly connected to color name responses than are color pathways. When Cohen, Dunbar, and McClelland trained the model in various ways, they were able to reproduce many of the major empirical results in the Stroop literature identified by MacLeod [5].

In 2003, Melara and Algom [17], coming from a fundamental perception perspective, proposed that two factors underlie Stroop interference: dimensional imbalance and dimensional uncertainty. Dimensional imbalance reflects how correlated the two dimensions of a stimulus are and how surprising a stimulus is and determines the ease of recovery of a stimulus representation from memory. Dimensional uncertainty reflects how salient a stimulus is, notably how likely or unlikely it is in the context of other (recently presented) stimuli. Together, these two factors determine the success of attentional selection by focusing on salient, surprising, and/or correlated information contained within each dimension and across the two dimensions of a Stroop stimulus. Each influences excitation of targets and inhibition of distractors. Stroop interference occurs both because there is more uncertainty in the colors than in the words and because the words are more salient than the colors. Their experiments manipulating these two factors were able to produce many of the major empirical results.

Also in 2003, Roelofs [18] proposed his model of Stroop interference, a model situated in an already implemented model of word production (WEAVER++) from the psycholinguistic literature. This also can be viewed as a two-factor model, with processing interactions occurring in the system that carries out language production, modulated by a supervisory attentional system that maintains task control. Roelofs posited that different architectures underlie color naming and word reading, with color naming, because it is conceptually driven, requiring an extra step due to colors not being directly connected to their names, unlike words. His experiments derived from this model also reproduced many of the major results in the Stroop literature.

The Big Picture

Anyone who has tried doing the Stroop experiment themselves knows that the interference caused by an incongruent word in naming its print color is powerful. As soon as we can read, we start to show this interference. Although we think of colors as being processed with ease, the existence of Stroop interference is an indication that there is computation involved in processing color names, a computation that can be disrupted. Some of this interference in the standard Stroop paradigm is response interference (i.e., the word and the color name are different responses within the small set), which reveals little about color processing. But the fact that color words not in the response set and words simply related to colors also produce interference demonstrates the semantic contribution to the Stroop effect.

This remarkable phenomenon has been used in recent years for many purposes, including to help to identify the function of certain brain regions, notably the anterior cingulate cortex and dorsolateral prefrontal cortex, both of which are active when resolving conflict. In the context of the Stroop task, the dorsolateral prefrontal cortex appears to be involved in the relevant executive functions, particularly in maintaining response set (to name the color, not to read the word), whereas the anterior cingulate cortex plays a central role in selecting the appropriate response and evaluating its accuracy [19, 20].

One thing is clear after 80 years: This simple and robust paradigm and the interference phenomenon that it demonstrates continue to provide valuable insights into the operation of cognitive processes. It would have been hard to imagine in 1935 that merely pitting a few colors against their names could provide such an important tool.

References

1. Stroop, J.R.: Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935)
2. Cattell, J.M.: The time it takes to see and name objects. *Mind*. **11**, 63–65 (1886)
3. Moors, A., De Houwer, J.: Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* **132**, 297–326 (2006)
4. Dyer, F.N.: The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Mem. Cogn.* **1**, 106–120 (1973)
5. MacLeod, C.M.: Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* **109**, 163–203 (1991)
6. MacLeod, C.M., MacDonald, P.A.: Inter-dimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends Cogn. Sci.* **4**, 383–391 (2000)
7. Williams, J.M.G., Mathews, A., MacLeod, C.: The emotional Stroop task and psychopathology. *Psychol. Bull.* **120**, 3–24 (1996)
8. Klein, G.S.: Semantic power measured through the interference of words with color-naming. *Am. J. Psychol.* **77**, 576–588 (1964)
9. Rosinski, R.R., Golinkoff, R.M., Kukish, K.S.: Automatic semantic processing in a picture-word interference task. *Child Dev.* **46**, 247–253 (1975)
10. Shor, R.E., Hatch, R.E., Hudson, L.J., Landrigan, D.T., Shaffer, H.J.: Effect of practice on a Stroop-like spatial directions task. *J. Exp. Psychol.* **94**, 168–172 (1972)
11. Windes, J.D.: Reaction time for numerical coding and naming of numerals. *J. Exp. Psychol.* **78**, 318–322 (1968)
12. Eriksen, B.A., Eriksen, C.W.: Effects of noise letters upon identification of a target letter in a non-search task. *Percept. Psychophys.* **16**, 143–149 (1974)
13. MacLeod, C.M., Dunbar, K.: Training and Stroop-like interference: evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 126–135 (1988)
14. Glaser, M.O., Glaser, W.R.: Time course analysis of the Stroop phenomenon. *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 875–894 (1982)
15. Dunbar, K.N., MacLeod, C.M.: A horse race of a different color: Stroop interference patterns with transformed words. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 622–639 (1984)
16. Cohen, J.D., Dunbar, K., McClelland, J.L.: On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* **97**, 332–361 (1990)
17. Melara, R.D., Algom, D.: Driven by information: a tectonic theory of Stroop effects. *Psychol. Rev.* **110**, 422–471 (2003)
18. Roelofs, A.: Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychol. Rev.* **110**, 88–125 (2003)
19. Banich, M.T., et al.: fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *J. Cogn. Neurosci.* **12**, 988–1000 (2000)
20. Milham, M.P., Banich, M.T., Claus, E.D., Cohen, N.J.: Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *NeuroImage*. **18**, 483–493 (2003)

Stroop Interference

► Stroop Effect

Structural Color

Richard J. D. Tilley
Queen's Buildings, Cardiff University,
Cardiff, UK

Definition

Structural colors arise when white light incident upon a material interacts with its structure such that some wavelengths (colors) are suppressed and others appear enhanced, without any significant light absorption occurring.

Description and Examples

White light is taken to mean unpolarized light from the sun or its equivalent. All of the visible wavelengths are present in the spectrum, from deep red, wavelength 700 nm, to deep violet, wavelength 400 nm. Structural colors, in the broadest sense, arise when white light interacts with a dielectric (usually transparent) material in such a way that the overall balance of this wavelength spread is disturbed *without* significant energy exchange between the incident beam and the constituents of the material, that is, no specific wavelengths are absorbed by the material during this encounter. The principal processes that give rise to such colors are scattering, refraction/dispersion, reflection, and diffraction. Frequently the structural color observed varies with the viewing angle, a phenomenon termed *iridescence*.

Structural colors were first characterized by Robert Hooke in his description of iridescent peacock feathers (*Micrographia*, 1665). He also gave a technique for distinguishing such structural colors from those that are due to pigmentation – when the color is wetted by an organic solvent, it

will disappear if its origin is structural but will be regenerated unchanged as the solvent evaporates. Examples of structural color abound in nature and include rainbows, halos, the colors of feathers, butterflies, beetles, and plants. The colored holograms used as security labels and for decoration are also structural colors, as are the colors that are seen when white light is reflected from compact discs.

Mechanisms

In the following descriptions, it should be noted that the mechanisms for the generation of structural colors described below are not always easily separated or distinct from each other. Many structural colors involve two or more of these mechanisms to generate the color effect, typified by the rainbow, which arises by a combination of dispersion and reflection.

Scattering

Scattering, which occurs when white light encounters small randomly located particles, gives rise to many colors that are found in nature, including blue skies, blue eyes, and the blue colors of some butterflies and feathers. Blue skies were first explained by Rayleigh, who investigated scattering by a small insulating (non-absorbing) sphere with a diameter less than one-tenth of the wavelength of the incident light. Scattering by such particles is referred to as *Rayleigh scattering*. In the case where a beam of unpolarized light is scattered once only by a single scattering center, the irradiance of the scattered light is proportional to $1/\lambda^4$, where λ is the wavelength of the light incident upon the particle. The blue color of the sky is a direct result of Rayleigh scattering by gas molecules in the atmosphere as the blue-violet region of the incident light is scattered approximately nine times more than red light, the result of which is that the sky away from the sun is perceived to be blue (Fig. 1). Towards sunset, when it is possible to look in the direction of the sun through a thicker layer of atmosphere, the scattering will remove blue light preferentially and the sun and sky will appear red.

Structural Color,**Fig. 1** Blue sky and white clouds, colors caused by light scattering

The irises of all eyes are blue at birth due to scattering. The iris is essentially a transparent material consisting of a composite of various tissues, small crystalline regions, and air vesicles, each of which have differing refractive indices. This inhomogeneity gives rise to preferential scattering of blue light. As the light transmitted through the lens and iris is absorbed in the underlying tissues, only the scattered light reemerges, to give the impression of blue. After some weeks, a pigment may be laid down in the iris and it is this deposited pigment that changes the color of the iris from blue to green or brown.

The “electric” clam, *Ctenoides ales*, has an “electric blue” display that is due to scattering from ambient light. The tissues responsible for this scattering contain a large number of silica spheres, amounting to a volume fraction of approximately 0.35, and of an average diameter of approximately $0.30\ \mu\text{m}$. This size is optimal for the scattering of blue and violet light when in water. The scattering layer is coupled with a strongly absorbing layer, so that the two layers, operating in concert, give rise to intense flashes of electric blue scattered light.

Scattering by larger insulating particles can be described by *Mie scattering* theory. Water droplets or ice crystals of varying sizes, but generally much larger than the wavelength of light, scatter all wavelengths approximately evenly: an effect

seen in clouds, mists, and fogs, which appear white (Fig. 1).

Dispersion

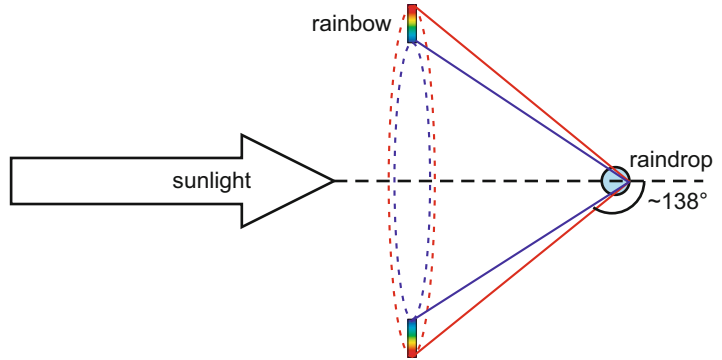
When a beam of white light enters a transparent material, each wavelength travels at a different speed, quantified as the *dispersion* of the material. Dispersion is measured by the refractive index, which is lowest for red light and greatest for violet light.

Because of dispersion, a beam of white light crossing a transparent material spreads out into a spectrum as the different wavelengths follow slightly different paths, with red deviated least and violet the most. This accounts for many colors, including the chromatic aberration of simple glass lenses and the “fire” of diamonds. It is the cause of rainbows, which are produced when sunlight coming from *behind* the observer illuminates a sheet of rain. Light entering each raindrop is dispersed and internally reflected, so as to approximately retrace its path towards the observer. A curtain of randomly oriented raindrops generates a cone-shaped spectrum (Fig. 2). The observer sees the upper part of this cone as an arc, with red outermost and violet innermost.

Halos, colored rings seen around the sun, or rarely the moon, are similar, and, unlike the rainbow, are seen on *looking towards* the sun or moon. These form when light passes through high thin

Structural Color,

Fig. 2 The formation of rainbow colors due to dispersion and reflection



Structural Color,

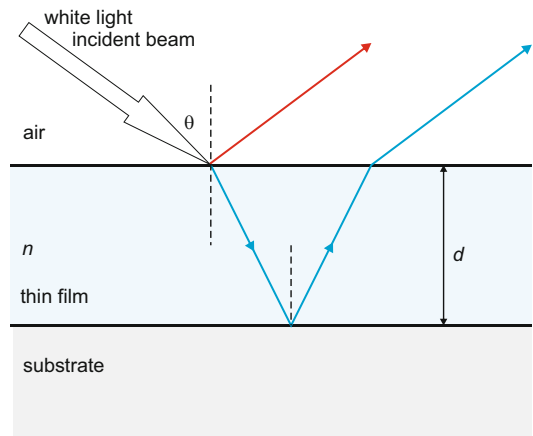
Fig. 3 Colors in a soap film caused by thin-film reflection



clouds that contain hexagonal ice crystals, which disperse the wavelengths to form visible colors. The most often observed halo subtends an angle of 22° to the observer's eye. As dispersion of the violet light is greater than the red, the circular halo is red internally and blue-violet externally.

Thin-Film Reflection

White light reflected from a transparent material in the form of a single thin film, either in air or on a substrate, can show vivid colors, typified by the colors seen in soap films and thin transparent insect wings (Fig. 3). This color arises because light waves are reflected from the upper surface and the lower surface of the film (Fig. 4). On leaving the film, these emergent waves now have a relative phase shift which is dependent upon the



Structural Color, Fig. 4 The production of thin-film colors, arising from the phase difference between rays reflected from the top (red) and bottom (blue) surfaces of a thin film

thickness of the film, its refractive index, and the wavelength of the light. When white light falls onto such a film, each wavelength will suffer a slightly different phase shift to its neighbors. This leads to interference, so that some wavelengths (i.e., colors) are canceled, while others are reinforced. The film then appears colored to an observer.

Although multiple reflections from both the top and lower surfaces occur, the production of color can be most easily understood if only that part of the incident beam that is reflected off the top surface and that part that is transmitted through the film and reflected once off the bottom surface are considered. At *normal incidence*, it is found that a monochromatic incident beam of wavelength λ reflected back from a thin film of refractive index n and thickness d in air or on a substrate of lower refractive index than the film (or in air) will show constructive interference and so appear *bright* when:

$$2nd = (m + \frac{1}{2})\lambda \quad (m = 0, 1, 2, 3 \dots)$$

When the light is incident at a small angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos\theta = (m + \frac{1}{2})\lambda$$

If the film is on a substrate of higher refractive index, the reflected light will show destructive interference and appear *dark* under the same conditions. Similarly, light reflected from a film on a substrate of lower refractive index than the film (or in air) will show destructive interference and so appear *dark* when:

$$2nd = m\lambda \quad (m = 1, 2, 3 \dots)$$

If the film is on a substrate of higher refractive index, the reflected light will show constructive interference and appear *bright* under the same conditions. When the light is incident at a small angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos\theta = m\lambda$$

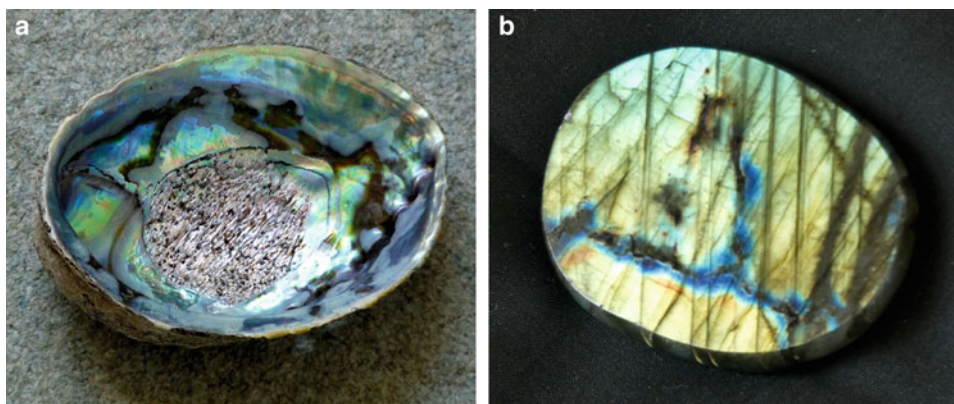
Note that the equations for viewing at an angle do not take polarization into account and must be employed with caution.

To determine the appearance of a film illuminated with white light, the reflectivity has to be summed across all of the wavelengths in the visible. The colors seen depend upon film thickness and cycle through several series or *orders*. The $\cos\theta$ angle dependence given above indicates that as the viewing angle moves away from perpendicular, the color observed moves towards shorter wavelengths: a *blue shift*. Because the color observed depends upon the film thickness and angle of observation, in the case of unstable films such as soap bubbles, the colors swirl and change as the film thins and flexes.

Many decorative films and paints are made by using thin-film iridescence. On bags and gift tags, this is often a thin polymer film, sometimes containing a dye, overlaying a reflective substrate. Metallic car paints often contain flakes of a mineral such as mica overlaid with a thin film of a transparent oxide such as TiO_2 .

A similar effect, but generally with greater brightness, is found if the reflecting object is composed of alternating lamellae of higher and lower refractive index. The multicolored mineral labradorite contains alternating lamellae of feldspars of slightly different constitution and refractive indices (Fig. 5a). The bright metallic-looking colors of many scarab beetles is due to a multilayer structure consisting of chitin and air layers, while in Paua shells, the color arises in layers of nacre, a composite of aragonite (CaCO_3) plates and proteins (Fig. 5b).

The reflectivity of such multilayers depends upon the thickness of the layers, their refractive indices, and the angle of view. Relatively simple formulae have been derived for the reflectivity of stacks where just two layer types are involved, each of which is $\lambda/4$ thick, (called a *quarter wave stack*), viewed at normal incidence. The peak reflectivity of such a stack with refractive indices n_1 and n_2 , of thicknesses d_1 and d_2 occurs at a wavelength given by $2(n_1d_1 + n_2d_2)$.



Structural Color, Fig. 5 (a) Paua shell; (b) Labradorite; both showing color due to multilayer reflection

There are a number of differences between the colors exhibited by multilayers compared to those from single thin films. In particular, the blue-shift observed as the observation angle moves away from normal to the stack is much less pronounced compared to single films.

Diffraction

When light encounters an object with dimensions similar to that of visible wavelengths, the waves are diffracted. The disturbance, called the diffraction pattern, is dependent upon the wavelength of the light and visible colors arise when the incident beam is white light. Two regimes have been explored in most detail, diffraction close to the diffracting object, giving rise to a *Fresnel diffraction* pattern, and far from the object, giving rise to a *Fraunhofer diffraction* pattern.

Isolated Centers

When light encounters a small hole, disc, speck or droplet, with a dimension similar to that of visible wavelengths, the diffracted beams take the form of nested cones with maxima and minima given by:

$$\sin\theta = m\lambda/d$$

where d is the diameter of the object and m takes values of: 0, the straight through beam; 1.220, first minimum (dark ring); 1.635, first bright ring, (called the first-order diffracted beam); 2.333, second minimum (second dark ring); 2.679,

second bright ring (second-order diffracted beam); and so on. For orders other than 0, the angle of scattering is wavelength-dependent so that each diffracted cone will consist of a spectrum, with violet diffracted least and red the most. These ring patterns are seen surrounding the image of a star when observed in a telescope, when they are called Airy's rings. Their appearance is indicative both of the quality of the optics of the system and of the atmosphere.

A dense random distribution of scattering centers of differing sizes will not show color as the orders overlap and cancel. However, if the collection is made up of similar-sized objects, colors can result. This occurs on rare occasions when high clouds are made up of uniformly sized droplets with diameters of the order of 0.003 mm and viewed at an angle of approximately 10–45° from the direction of the sun, a phenomenon called *irisation*. Closely related are the coronae (colored rings) that are sometimes seen by an observer looking towards the source of light, the sun or moon, through mist or thin clouds. They are narrower than halos, usually 1–10° in diameter and surround a white central area called the aureole. The pattern is the Fraunhofer diffraction pattern arising from the randomly distributed droplets in the mist layer. These add together and an observer will see fragments of the diffraction patterns from many droplets or specks, each of which contributes to the overall effect. When the clouds consist of similarly sized droplets or specks, the effect will be strong. Coronae always

show the sequence of colors violet-blue, green, yellow, red, with red outermost.

A multicolored corona-like ring due to diffraction can sometimes be seen to surround a narrow beam white light which has passed through a pane of glass covered with a fine powder or is reflected from a mirror covered with fine dust or droplets.

Sheets

A surface which contains a series of equally spaced parallel grooves or slits with a spacing of about the wavelength of visible light forms a *one-dimensional diffraction grating*. Light incident upon such a grating of spacing d (either in transmission or reflection) will be scattered strongly in directions given by the *grating equation*:

$$d (\sin\theta_i + \sin\theta_m) = m\lambda$$

where θ_i and θ_m are the angles of incidence and diffraction, respectively, of the order m , and θ_m is positive when the diffracted and incident beam lie on the same side of the grating normal and negative when they lie on opposite sides of the normal. The Fraunhofer diffraction pattern will consist of a row of bright spots called *principal maxima*. The spacing of these spots is proportional to $1/d$ and lie along a line perpendicular to the grating grooves. They are separated by much smaller secondary maxima which do not give rise to strong color effects. The zero order, $m = 0$, diffracted beam is equivalent to an undeviated beam (transmission) or specular reflection (reflection). For the other orders, the angle of scattering, θ_m is wavelength dependent so that white light will be spread out into a spectrum, with violet diffracted least and red the most.

Scattering centers, such as pits, protuberances, or apertures distributed across a surface to form a regular lattice with a spacing of the order of the wavelength of light, form a *two-dimensional grating*. Each row on the pattern acts as a one-dimensional grating, and the diffraction pattern may be understood intuitively as arising from intersecting one-dimensional gratings. Each row will give rise to principal maxima, (described by the grating equation) running normal to the row selected, separated by weak intermediary

subsidiary spots that do not give rise to strong colors. A complete pattern will only be seen if the conditions of the grating equation are fulfilled for every row simultaneously. For a rectangular grating illuminated by a beam normal to the grating, the conditions for diffraction are:

$$a \sin\theta_m = m \lambda \text{ (grating spacing } a)$$

$$b \sin\theta_p = p \lambda \text{ (grating spacing } b)$$

Each diffracted order is specified by two indices m and p , so that:

$$d_{mp} \sin\theta_{mp} = mp \lambda$$

where d_{mp} is the spacing of the grating elements along the row giving rise to the diffraction pattern row containing the order mp . As before, the zero-order beam corresponds to undeviated transmission or specular reflection and is not colored. The diffraction orders, each forming a spectrum with violet diffracted least and red the most, will now be arranged on a grid with symmetry that matches that of the grating.

Many iridescent gift tags and bags show two-dimensional diffraction effects. The surface of these consists of a thin plastic film, often overlaying a reflecting layer. The polymer surface is embossed with several sets of grooves. Two sets at right angles give rise to square or rectangular diffraction effects while three sets at 60° give rise to hexagons and star diffraction patterns. Compact discs, which store information by way of ordered tracks of surface pits, similarly act as reflection diffraction gratings and show colors when viewed at the correct angle in white light (Fig. 6).

Three-Dimensional Arrays

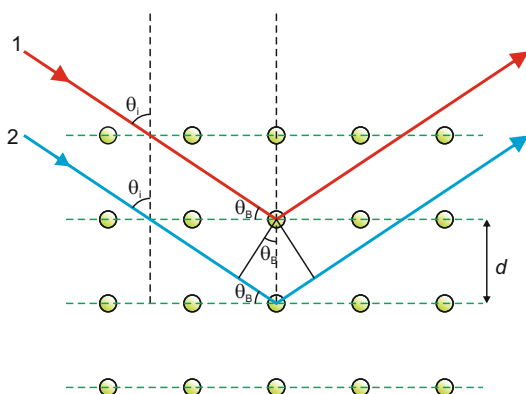
Photonic Crystals

The theory of diffraction from crystals, ordered three-dimensional arrays of scattering centers, was first derived with respect to X-ray diffraction. The condition for strong diffraction to occur, Bragg's law, is:

$$m\lambda = 2d \sin\theta_B$$



Structural Color, Fig. 6 Colors generated by a reflection diffraction grating on the surface of a compact disc



Structural Color, Fig. 7 Bragg's law: diffraction by a three-dimensional array of scattering centers

where m is the order of the diffracted beam, (1, 2 ...), d is the distance between the planes of scattering centers giving rise to the diffracted beam, and θ_B is the angle between the incident beam and the scattering planes (Fig. 7).

Three-dimensional ordered crystal-like arrays of scattering centers built from insulating materials with different refractive indices, and dimensions approximately equal to the wavelength of light are a well-known source of intense diffraction color and are now referred to as *photonic crystals*. These materials, also called inverse opals or colloidal photonic crystals, can be made

up of arrays of particles, voids, tubes, rods, and so on, all stacked in a regular crystal-lattice like arrangement.

The behavior of photonic crystals can be illustrated by reference to natural opals. These gemstones consist of arrays of ordered silica spheres embedded in a matrix of amorphous silica. When an opal is illuminated with white light, it strongly diffracts various colors. As the opal is tilted, any particular observed color moves to shorter wavelengths giving a blue shift. To quantify this, Bragg's law must be modified to include the effective refractive index of the opal:

$$m \lambda = 2 d \sqrt{[n_e^2 - \sin^2 \theta_i]}$$

where m is the order of the diffracted beam, d is the spacing between the layers of spheres or voids that make up the diffracting planes of the crystal, and θ_i is the (conventional) angle of incidence of the white light in medium 1 with the *effective refractive index* of n_e . The position of all diffracted beams except for the zero order is a function of wavelength and each diffracted order will be pulled out into a spectrum with violet diffracted least and red the most. As the angle of incidence increases, the wavelength diffracted will decrease, that is, a red reflection at $\theta = 0^\circ$ will move towards green and blue, as described above.

Many beautiful colors in nature are produced by natural photonic crystals created by living organisms. Probably best known are some of the spectacular colors of certain butterflies or beetles, but lesser known creatures, such as the sea mouse, have vividly colored spines due to a photonic crystal. Here just one example is given to illustrate this phenomenon. The green hairstreak, *Callophrys rubi* has the underwings colored a mat green (Fig. 8a). The color arises in the layer of scales that cloak these surfaces. The internal structure of each scale consists of a perforated membrane in the form of a photonic crystal (Fig. 8b), which produces the green color observed.

Cholesteric Liquid Crystals

Cholesteric liquid crystal mesophases are built up of layers that are composed of birefringent



Structural Color, Fig. 8 (a) The green hairstreak butterfly, *Callophrys rubi*; (b) a green scale, showing the internal structure. [(a) Courtesy Dr. J. A. Findlay]

molecules. The simplest are the *chiral nematic phases*. In these phases, the director (the average optical axis of the slightly disordered molecules) rotates as one travels in a direction perpendicular to the sheets of molecules (Fig. 9). The physical reason for the generation of a helical structure is that the molecules in each layer are asymmetric. When such molecules pack together, the interactions are minimized if molecules in one layer rotate slightly compared to those in the layer below so as to reduce intramolecular interactions. Because these interactions are the same from one layer to another, the same twist occurs between each layer. In this way, a uniform helical structure results. The defining scale in these liquid crystals is the distance over which the director rotates by 360° , which is called the *pitch* (p) of the helix. The pitch of the helix can be modified by attaching extra groups to the parent molecule of the structure, a feature widely used in the fabrication of liquid crystal-based devices.

White light incident upon such a layer structure will be colored in the same way as photonic crystals. The color observed is given by Bragg's law, in which the pitch of the spiral is equivalent to d and the observed color changes with viewing angle as in the case of opals. The conditions for this to occur follow Bragg's law, which, for cholesteric liquid crystals, is more often written as:

$$\lambda = 2n_m p \cos \theta$$

where n_m is the average refractive index of the mesophase molecules:

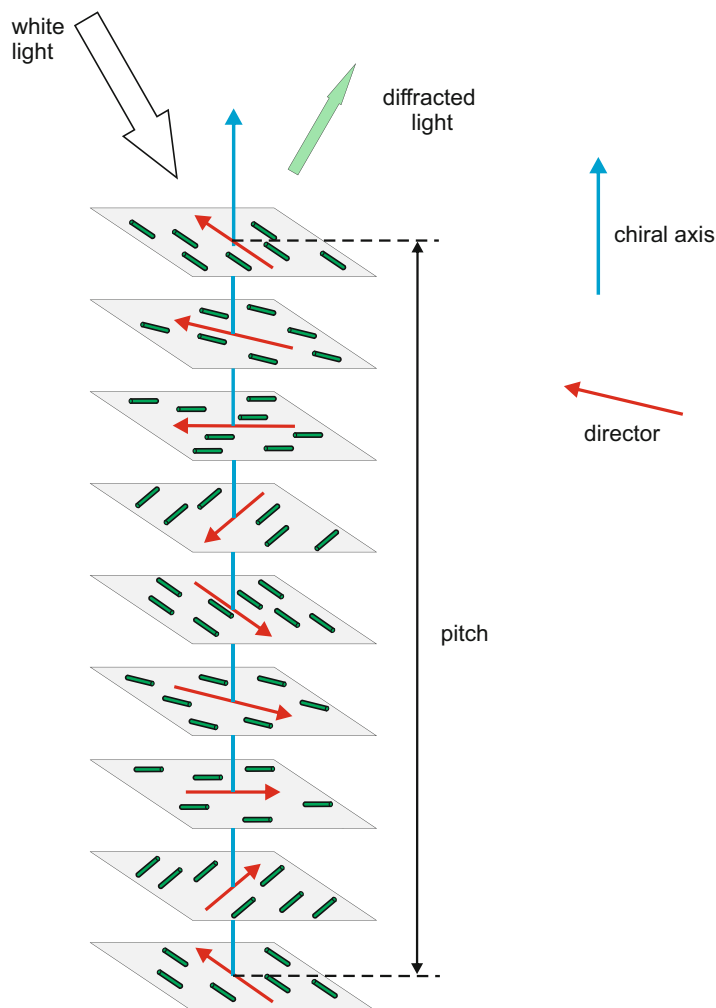
$$n_m = (n_e + n_o)/2$$

and θ is the angle between the incident light direction and the axis of the helix, that is, the normal angle of incidence used in optics.

The helix formed by the molecular stacking can be right- or left-handed, and this influences the polarization of the diffracted beams. Normal white light can be regarded as two beams of oppositely circularly polarized light. When the light is normally incident upon the helix, the beam with the same handedness as the helical arrangement of the mesophase is diffracted (reflected) while the component with the opposite handedness is transmitted. When light falls at an oblique angle to the helical axis, both the transmitted and reflected light will be elliptically polarized.

Structural color showing strong polarization is found in a number of beetles. This color effect was first clarified by Michelson in 1911 for the scarab beetle *Chrysina (Plusiotis) resplendens*. In this insect, the elytra, the hard wing cases that cover the back of the beetle, show a metallic green iridescence. Other related beetles with a similar metallic sheen are *Chrysina (Plusiotis) gloriosa* (green/gold), *Chrysina (Plusiotis) optima* (green/blue), and *Anomala dimidata* (green). The light reflected from these last three insects is left

Structural Color,
Fig. 9 Diffraction from a
 cholesteric chiral nematic
 liquid crystal structure
 (schematic)

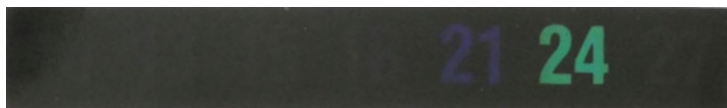


circularly polarized, while that from *C. resplendens* shows both left and right circularly polarized components.

A number of plants also exhibit helicoidal structures and reflect metallic/iridescent circularly polarized light. Here the structures consist of helicoidally arranged layers of cellulose fibrils in a noncrystalline matrix. The metallic blue-green fruits of the plants *Margaritaria nobilis* and *Pollia condensate* are colored in this way. The young leaves of the ferns *Danaea nodossa*, *Diplazium crenatosseratum*, *Diplazium tomentosum*, and *Lindsaea lucida*, all show metallic blue iridescence. Similarly, the flowering plant *Phyllagathis*

rotundifolia and the sedge *Mapania caudate* also show metallic colors that arise in helicoidal structures.

The pitch of the helical structure can be modulated by external stimuli, including temperature, a property used in liquid crystal thermometers (Fig. 10). These devices operate in the following way. A series of spots of a cholesteric material are arranged in a row. They are chosen so that the periodicity of the molecular helix in each spot will diffract visible light at a precisely defined temperature. An increment of the temperature, δT , of the order of 1 °C, is engineered between each successive spot by using additives or by slightly

Structural Color,**Fig. 10** A liquid crystal thermometer strip

modifying the cholesteric molecule. Within the design temperature range, each spot will diffract visible light only when the mesophase is at the correct temperature, which causes the appropriate temperature value to shine. On approaching this temperature, each spot will run through a spectrum of colors as the pitch of the helix varies.

Rainbow Holograms

Rainbow or Benton holograms are the brightly colored holograms seen on credit cards, security documents, and so on. The hologram itself consists of a transparent photopolymer film with a reflective backing. To view these holograms, white light passes through the transparent layer and is reflected from the backing layer before returning to the observer. The hologram film imposes information on the incident white light by altering its phase, by different amounts, across the incident beam. This is achieved by a spatial variation in the refractive index of the hologram film. Although this refractive index variation is three-dimensional in nature, it is not ordered, as in a photonic-crystal lattice, but, instead, preserves the complexity of the intersecting wave fronts that formed the hologram. The colors that are seen by the viewer, the reconstructed image, are then produced as a result of interference of the phase modified beams after traversing the film.

The information encoded in the hologram is, in essence, an image of the object seen through a narrow slit-shaped window. Each wavelength present in the white light will create its own window, with red corresponding to most diffracted light and violet the least. Thus an observer on a level with the hologram will see a green-dominated reconstruction, while if the observer moves up, the image color will shift through yellow and orange towards red, or, if the observer moves down, will shift to blue, indigo, and violet. There will still be considerable overlap of colors in parts of the image, dependent upon the viewing distance and angle, so that a multiplicity of colors

and patterns can be picked out as the hologram is tilted and moved.

Cross-References

► [Iridescence \(Goniochromism\)](#)

References

1. Isenberg, C.: *The Science of Soap Films and Bubbles*. Dover. Dover Publications Inc. New York (1992)
2. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Lynch, D.K., Livingston, W.: *Colour and Light in Nature*. Cambridge University Press, Cambridge (1995)
5. Meadows, M.G., Butler, M.W., Morehouse, N.I., Taylor, L.A., Toomey, M.B., McGraw, K.J., Rutowski, R. L.: Iridescence: views from many angles. *J. Roy. Soc. Interface.* **6**(Suppl_2), S107–S265 (2009)
6. Vukusic, P.: Structural colour. In: Schwartz, J.A., Contescu, C.I., Putyera, K. (eds.) *Dekker Encyclopedia of Nanoscience and Nanotechnology*, vol. 5, pp. 3713–3722. Marcel Dekker, New York (2004)
7. Lee, D.: Chapter 10, Iridescent plants. In: Lee, D. (ed.) *Nature's Palette*. University of Chicago Press, Chicago (2007)
8. Parker, A.R.: 515 million years of structural colour. *J. Opt. A Pure Appl. Opt.* **2**, R15–R28 (2000)
9. Vukusic, P., Sambles, J.R.: Photonic structures in butterflies. *Nature*. **424**, 852–855 (2003)
10. Kinoshita, S., Yoshioka, S., Miyazaki, J.: Physics of structural colours. *Rep. Prog. Phys.* **71**, 076401 (2008)
11. Sun, J., Bhushan, B.: Structure and mechanical properties of beetle wings. *RSC Adv.* **2**, 12606–12623 (2012)
12. Sun, J., Bhushan, B., Tong, J.: Structure colour in nature. *RSC Adv.* **3**, 14862–14889 (2013)
13. Cuthill, I.C., et al.: The biology of colour. *Science*. **357**, 470 (2017)
14. Burg, S.L., Parnell, A.J.: Self-assembling structural colour in nature. *J. Phys. Condens. Matter.* **30**, 413001 (2018)
15. Kawamura, A., et al.: Full-colour biomimetic photonic materials with iridescent and non-iridescent structural colours. *Sci. Rep.* **6**, 33984 (2016)

Structure Property Relationships

- [Colorant, Environmental Aspects](#)

Subjective Colors

- [Fechner's Colors and Benham's Top](#)

Subsampling

- [Interpolation of Spectral Data](#)

Surface Plasmon Polariton

- [Surface Plasmons](#)

Surface Plasmon Resonance

- [Surface Plasmons](#)

Surface Plasmons

Li-Lin Tay
Measurement Science and Standards, National
Research Council Canada, Ottawa, ON, Canada

Synonyms

[Plasmonics](#); [Surface plasmon polariton](#); [Surface plasmon resonance](#)

Definitions

Surface plasmons are collective excitations of conduction electrons at the interface between a

metal and dielectric that are stimulated by electromagnetic radiation.

A *surface plasmon polariton* (SPP) occurs when surface plasmons interact strongly with electromagnetic radiation.

Surface plasmon resonance (SPR) is another synonym for SPP and refers to the coherent (resonant) oscillation of the surface conduction electrons excited by electromagnetic radiation.

A *localized surface plasmon resonance* (LSPR) is a nonpropagating SPP confined to nanostructured surfaces.

The term *plasmonics* is often used to cover the various phenomena sustained by SPP and SPR types of light-matter interactions.

Overview

The electron charge density and its electromagnetic fields propagate as a surface wave along a metal–dielectric interface. The electromagnetic field intensity decays exponentially away from the interface. Material systems with free or almost free electrons will sustain SPP excitation. More generally, an SPP occurs when the real part of the dielectric function of the metal is negative and its magnitude is larger than that of the complex dielectric function. This also means that an SPP can only exist for transverse magnetic polarization since no surface modes exist at the metal surface for transverse electric polarization. For common noble metals, such as gold and aluminum, this condition is met in the visible–NIR wavelength region at the air–metal interface. As an SPP propagates along the metal–dielectric boundary, it is very sensitive to changes that modulate this interface. SPP can be excited by electrons or photons. A planar surface of a good conductor that sustains SPP with given frequency ω and parallel momentum obeys the following dispersion relation [1]:

$$k_{\parallel}^2 = \left(\frac{\omega}{c}\right)^2 \operatorname{Re} \left[\frac{\varepsilon_0 \varepsilon}{\varepsilon_0 + \varepsilon} \right]$$

where $k_{\parallel}$ is the parallel component of the wavevector, $\varepsilon(\omega)$ is the dielectric function of the conductor, c is the speed of light in vacuum, and ε_0

is the dielectric constant of the ambient medium (air or vacuum), also known as the permittivity.

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}$$

where bulk plasma frequency ω_p is

$$\omega_p = \sqrt{\frac{ne^2}{\epsilon m_{\text{eff}}}}$$

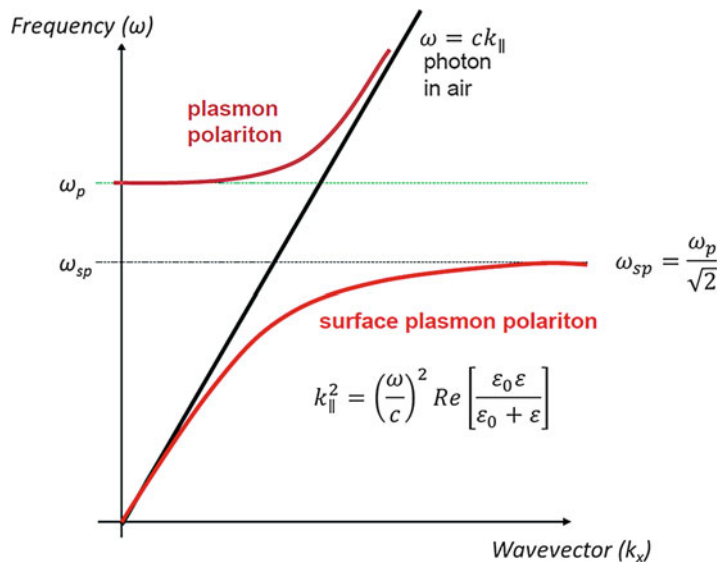
with n and e being the free electron density and electron charge, respectively, and m_{eff} the electron effective mass; ϵ here is the vacuum permittivity. Figure 1 depicts the SPP dispersion relation. At lower wavenumber (mid-IR or lower frequency), SPP exhibits photon characteristics, and the waves extend over many wavelengths into the dielectric space. In this regime, SPP behaves like a grazing-incidence light field. As the wave vector k increases, the dispersion relation tends toward the asymptotic limit of the surface plasma frequency, ω_{sp} . Due to their bound nature, the SPP dispersion relation lies entirely to the right of the light line (photon in air, with $\varepsilon_0 = 1$). In this regime, the out-of-plane component of the SPP wave vector ($k_{\perp}$) is purely imaginary and decays evanescently away from the metal surface. Thus, special phase-matching

techniques such as grating or prism coupling are required to excite SPP.

For an incident electromagnetic wave to excite the SPP, both the frequency and parallel momentum must be conserved. This condition cannot be achieved with air (or vacuum) as the ambient medium at the interface due to wave vector mismatch. For the same reason, an SPP from a smooth flat surface does not radiate into the ambient. It is confined to the metal surface with its amplitude decaying exponentially away from the metal surface. SPP propagates on the metal surface and eventually dissipates its energy as heat through Ohmic loss and electron core interaction. For a relatively absorbing metal such as aluminum, SPP propagation lengths can reach 2 μm when excited with 500 nm (visible, green) radiation. For a low-loss metal such as silver where there is less damping, excitation at the same wavelength increases the SPP propagation length to 20 μm . Moving to a longer wavelength where there is less confinement, for example an excitation wavelength of 1.55 μm typically used for optical fiber-based telecommunication, the SPP propagation length can increase into the mm range. The propagation length of SPP sets the upper size limit for the SPP-based photonic circuitry. However, its strength lies in its ability to transmit information in subdiffraction limited

Surface Plasmons,

Fig. 1 Dispersion relation of plasmon polariton and surface plasmon polariton



circuits as shown in the section on examples and applications below.

Examples and Applications

Gold, silver, copper, and aluminum are among the many metals that exhibit strong plasmonic responses. These metals have long been known to have optical properties distinct from standard dielectrics. In fact, long before scientists began to investigate the optical properties of metal nanostructures, plasmonic nanoparticulates were used to produce brilliant colors in ancient artifacts. One of the best-known examples is the famous Lycurgus cup (Fig. 2) that is on exhibit at the British Museum. This ancient Roman relic was constructed with one of the few dichroic glasses that appear opaque and green in reflected light and a brilliant translucent red in transmitted light. The difference in color is due to the distinct plasmonic properties of colloidal gold and silver which are present in trace amounts in the glass of the cup. Studies from a small fragment of the relic have revealed trace amounts of 20–30 nm colloidal gold and silver nanoparticles which give the glass its unique dichroic properties. Similar inclusions of colloidal metal nanoparticles have been found in brilliantly colored stained glass windows in churches.

Surface plasmon resonance (SPR), also known as SPS, is a popular analytical tool for the study of

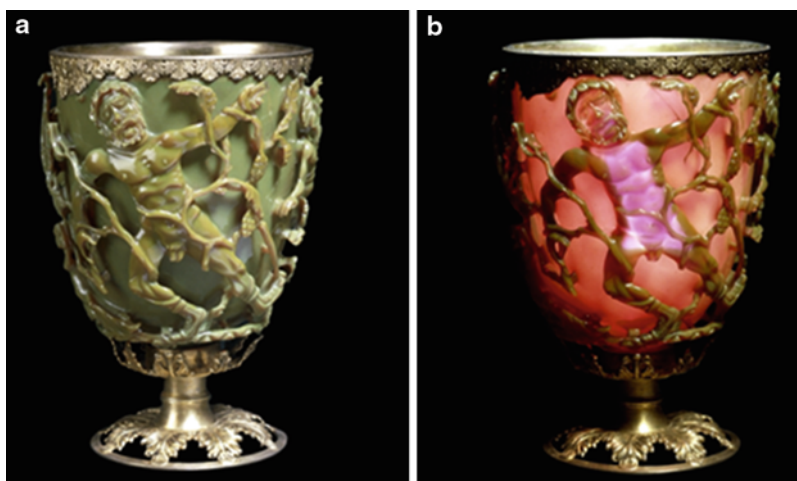
biomolecular interaction, the determination of kinetic reaction rates, and the evaluation of binding affinity of measured molecular binding events.

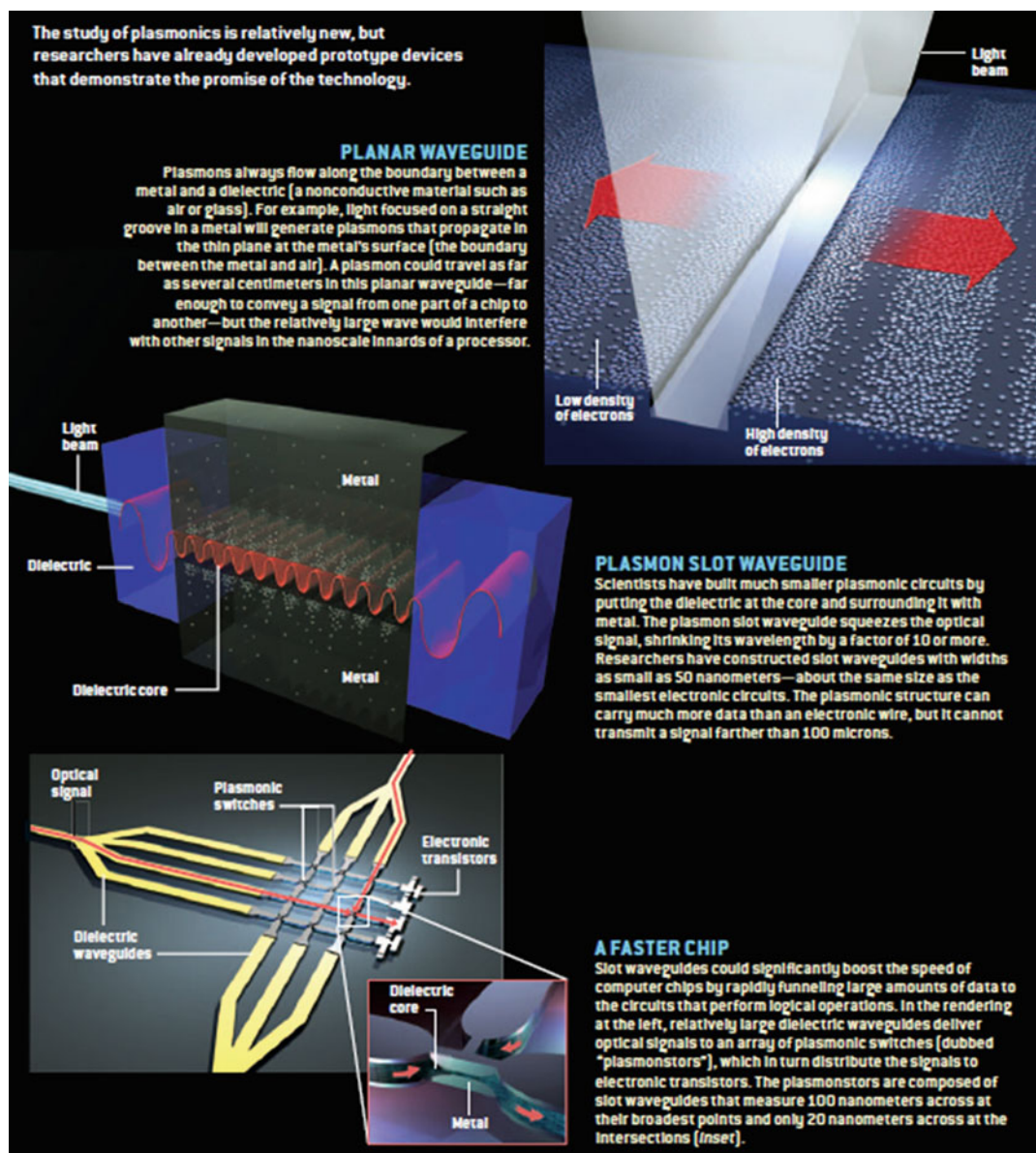
More recently, applications of SPP involve the design and fabrication of metal nanostructures to tune their plasmonic properties. This requires delicate chemical synthesis with protocols designed to disperse metallic nanostructures tagged with desired molecular recognition units or nanofabrication of specifically structured metal nanostructures to enable the confinement and/or propagation of SPP to transmit information. While it may seem impractical to transmit light signals with metal structures, the confinement of SPP to the metallic nanostructures enables the propagation of information in subdiffraction limited circuitry. Because planar plasmonic structures act as waveguides confining SPP along the metal–dielectric boundary, they can be used to route signals on a chip. Metal circuits can be defined lithographically and mass produced as miniature-sized plasmonic devices enabling very high density and fast chips as shown in Fig. 3 below [2].

Another striking application of SPP comes in the form of the localized surface plasmon resonance (LSPR) which is essentially SPP that is confined to small metal nanostructures, most often in the form of metal nanoparticle aggregates. Excitation of an LSPR can be tuned by material composition, size, shape and coupling between nanostructured entities. Similar to SPR, LSPR is widely used in biosensors for the study of small

Surface Plasmons,

Fig. 2 The Lycurgus cup from the Roman Empire, fourth century AD. The cup displays a *green colour* (a) when viewed with reflected light and a brilliant *red colour* when viewed with transmitted light (b)





Surface Plasmons, Fig. 3 Examples of SPP waveguides and an example of plasmonic circuitry (Sources: Scientific America) [2]

molecule interactions. Localization of electromagnetic fields at certain regions of the nanostructure due to resonant excitation of SPP is the foundation of many forms of surface enhanced spectroscopy and colorimetry based bioassay.

The detailed mechanism of LSPR is described in the section on surface enhanced Raman

spectroscopy below, as it is the dominant mechanism that enables this form of spectroscopy. Here we will simply describe the photothermal effect and its theranostic application. Gold nanoparticles (NPs) are one of the most studied nanomaterials for applications in the biomedical sciences. Gold NPs are potential drug carriers, photothermal

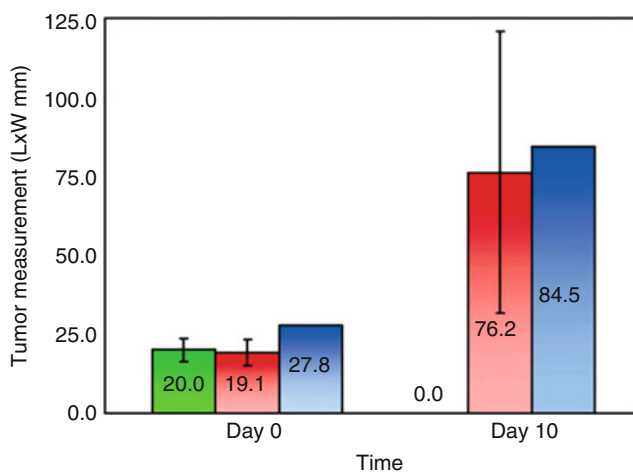
treatment agents, radiosensitizers, optical imaging contrast agents and have shown great promise for cancer therapy. There are a few different types of gold NPs: solid Au NPs, Au nanoshells, and Au nanorods that are commonly used in biomedicine. These different types of NPs are differentiated by their characteristic localized surface plasmon resonances. Among them, Au nanoshell and Au nanorods are known to exhibit broad LSPR tunability, covering much of the visible and NIR spectral window. Synthetic protocols and surface functionalization methods are well publicized through peer reviewed scientific literatures [3]. The nanoparticles are typically stabilized and guided to the targeted tissues by coating them with polyethylene glycol and antibodies. As mentioned in the earlier discussion, surface plasmon resonance interaction with metal nanostructures is a lossy process with most of its energy being dissipated as heat through Ohmic loss and electron-core interaction. The heat generated through optical excitation of LSPR provides gold NPs with a unique photophysical property to act as an anti-cancer agent. Tumor-specific ligands (e.g., transferrin, folic acid, antibodies) have been attached to the surface of gold nanoparticles for specific targeting of and delivery to tumor sites.

There are currently a few different versions of gold-NP based cancer treatment therapy that are undergoing various phases of FDA (Food and Drug Administration) approval [4]. One version

of such anti-cancer treatment relies on the plasmonic properties of a gold nanoshell. The gold nanoshell can be engineered to resonantly interact with different wavelengths of the light used to excite the LSPR. This is done by tuning the shell thickness of the gold coating over the silica nanosphere. With the LSPR of Au nanoshell tuned to a near-infrared absorption window, it can be delivered to cancer sites with the help of targeting agents as mentioned above. Upon arriving at the tumor site, a near infrared laser is used to excite the Au nanoshell and trigger the photothermal effect. This turns the optical radiation into intense but localized heating sources thereby enabling selective and rapid tumor destruction with minimal damage to surrounding tissues. Photothermal therapy of head and neck cancer is currently under phase I of FDA approval. Phase I of clinical trial is aimed to evaluate safety, determine safe dosage range and identify side-effects of new drugs. Photothermal therapy may also be used in combination with standard chemotherapy and radiation. This concept was pioneered by Professors Naomi Halas and Jennifer West of Rice University in 2004 [5]. Figure 4 below shows therapeutic efficacy and tumor growth over a period of time. The figure showed the original mean tumor size at the beginning of the treatment (day 0) and day 10 of the photothermal therapy. The treatment group (with Au nanoshell injected to tumor site) is shown in green, control group (receives no treatments) in red and sham treatment

Surface Plasmons,

Fig. 4 Animal study showing gold nanoshell efficacy in cancer treatments [5]. Mean tumor size at the beginning of the treatment and on day 10 for the treatment group (*green*), control group (*red*), and sham treatment (*blue*, no nanoshell but treated with same dose of NIR irradiation)



group (which receives same laser treatment as treatment group with no nanoparticle) in blue.

Surface Enhanced Spectroscopies

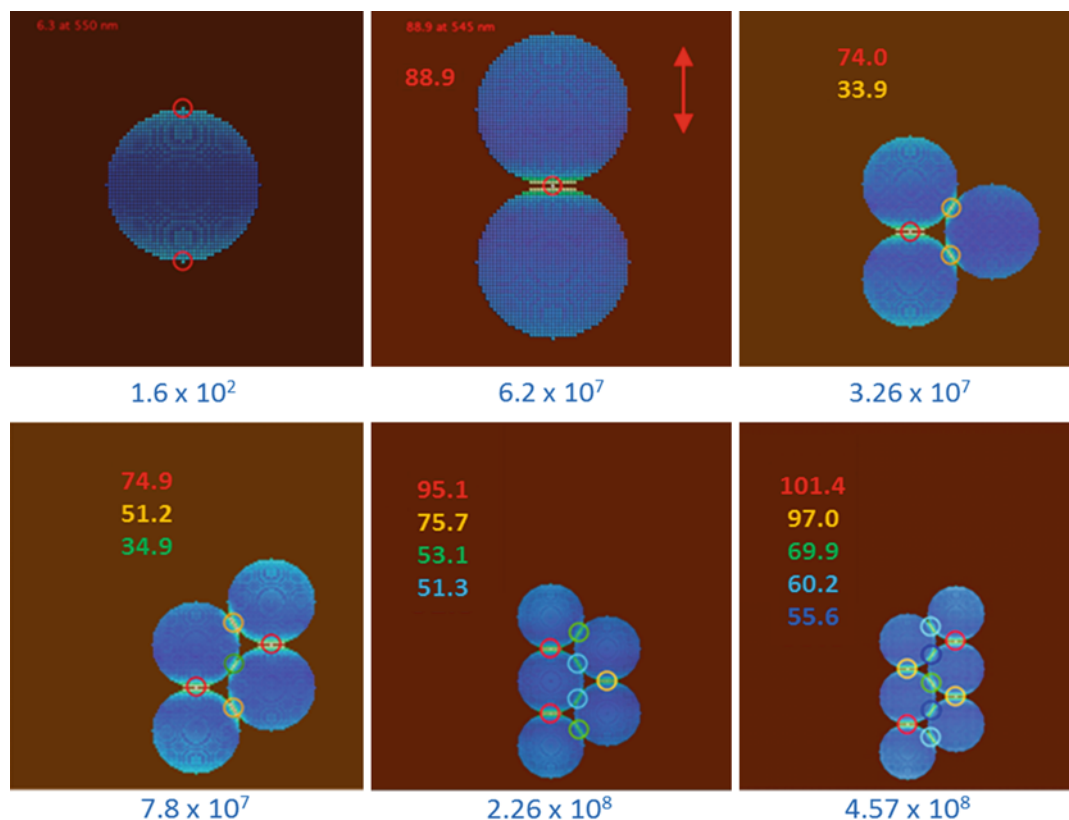
Raman spectroscopy is a vibrational technique which provides a wealth of information on the molecular species under investigation. Its many fields of application include identification and structural characterization of semiconductor materials, authentication of artifacts and legal documents, quality control in pharmaceutical manufacturing and the development of new biophysical tools for disease diagnosis and health monitoring. A major disadvantage of Raman spectroscopy is the inherently small scattering cross-sections of the Raman process, which are typically 12–14 orders of magnitude smaller than fluorescence cross sections. This was why the observation of an unexpectedly large Raman signal from pyridine adsorbed on roughened Ag electrode, reported in 1974 by Fleischmann et al., surprised researchers in the field [6]. The Ag electrode had been roughened through repeated oxidation-reduction cycles in an attempt to develop an in situ monitoring tool that is chemically specific [6]. The authors attributed the large pyridine Raman signal to the increase in number of molecules adsorbed on the increased surface area of the roughened Ag electrode. In 1977, Van Duyne et al. recognized that the large observed intensity could not be accounted for by the increase in the number of scatterers and proposed an electric field enhancement model [7]. Ultimately, it was the 1978 analysis by Moskovits [8] which attributed the predominant contributing factor of SERS (surface Enhanced Raman Scattering) to be the electromagnetic excitation of localized surface plasmon resonance (SPR) sustained by properly designed metal nanostructures. His interpretation cements the electromagnetic enhancement theory for SERS and paved the way for the development of the Plasmonics field. It is important to also note that LSPR is not the sole contributing factor to the inordinate large Raman intensity observed from molecules associated with metal nanostructures.

An alternate theory proposed by Albrecht and Creighton in their 1977 report proposed that resonance Raman scattering from molecular electronic states, broadened by their interaction with metal surface, could be responsible for the increased intensity [9]. It is understood that both factors contribute to the observed SERS signature with the LSPR being the dominating enhancement mechanism.

The plasmonic interpretation of SERS led to the understanding and prediction of many characteristic SERS effects. One of them is the fourth power dependency of the SERS enhancement factor (G) upon the local incident electric field ($|E_0|$); $G = |E_0|^4$. Another well-known characteristic of the plasmonic nature of SERS is the coupling of LSPR between adjacent nanostructures which results in extraordinary local field strength thereby enabling single molecule detection from tightly coupled nanostructures. Figure 5 below shows the false coloured electric field distribution of a single gold nanoparticle and a variety of coupled gold nanostructures.

LSPR is extremely sensitive to composition, morphology, geometry and degrees of aggregation and dielectric properties of the ambient environment as partly demonstrated in Fig. 5 [10]. All these factors, while affecting the overall strength of the observed SERS signal, also enable flexibility in the design of nanostructures to sustain stable SERS.

SERS is a powerful spectroscopic effect that leverages the enormous electromagnetic field enhancement caused by the excitation of an intense, sharp and localized surface plasmon resonance of metal nanostructures. The effect can provide rich structural and chemical information for molecules adsorbed on or bonded to SERS active nanometer-sized metal structures. Recent advances in nanotechnology and nanofabrication techniques have enabled researchers to design SERS active colloidal NP systems with optimized electromagnetic enhancements engendering a large variety of SERS-based bioanalytical and bioimaging applications. The advantages of SERS over standard fluorescent labeling techniques lie not only in its multiplexing capability and superior photostability, but also in the



Surface Plasmons, Fig. 5 Electric field distribution in different 2D geometric gold nanoclusters

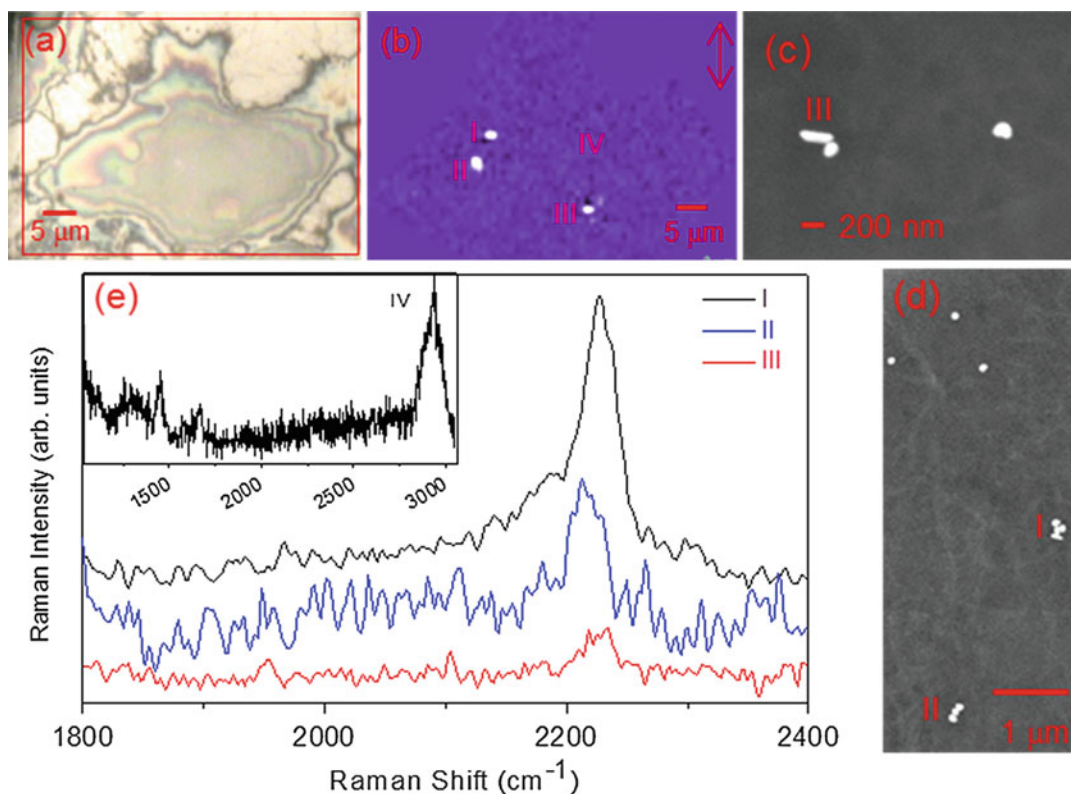
potential multimodal functionality provided by the other intrinsic properties of the metal NPs, such as the photothermal response which enables its therapeutic role in the treatment of cancer as outlined earlier [11]. In addition to its cancer fighting ability, the intense and unique signal from SERS nanoparticles also enables diagnostic imaging. SERS labels rival that of fluorescent contrasting agents and can be adapted to most of the fluorescent detection schemes. Figure 6 below shows an example SERS imaging of cell surface receptors in mammalian cells [12]. By the similar targeting mechanism, nanoparticles can be guided towards diseased tissue sites and optical spectroscopy-microscopy can be used as a diagnostic means to pin point the disease sites.

In addition to biomedical applications, SERS is also very popular in security applications and

trace chemical detection owing to its extraordinary sensitivity and information rich spectra.

SPR Sensors

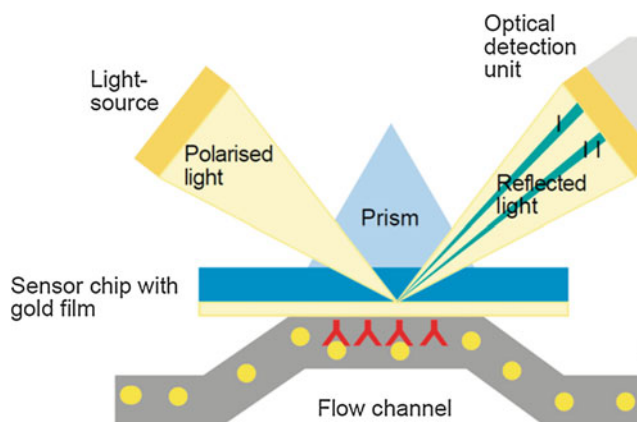
The surface plasmon resonance (SPR) enables many forms of enhanced spectroscopy but the simplest form of SPR sensor relies on very simple SPR reflectivity to detect molecular absorption, binding dynamics such as association and dissociation constants of a multitude of molecules from simple organic molecules, polymers, and DNA to complex proteins and antibodies. In its most common form, a flat thin Au film is deposited or index matched to a prism as shown in Fig. 7 below. Users measure the angle of minimum reflection which corresponds to the maximum absorption of



Surface Plasmons, Fig. 6 (a) Optical image of a HeLa cell labeled with nitrile-reporter and target-specific molecules. (b) SERS intensity map of the nitrile vibrational band with three SERS hot spots I, II and III identified. (c) and (d) SEM (scanning electron microscopy) image of

the corresponding area showed SERS activity. (e) SERS spectra of the three spots identified in the intensity map. *Inset* shows the native cell Raman spectrum. *Inset* of (e) shows the cell Raman spectrum taken from the nucleus region labeled as spot IV

Surface Plasmons, Fig. 7 Typical SPR sensor setup



the Au film. The reflection angle changes as film thickness changes. This is due to the adsorbed molecules modulating the effective index of

refraction of the film/adsorbate system and hence the plasmon resonance condition. The SPR sensors on the Au film can be pre-functionalized with

a capture molecule. When the molecule of interest is present in the test solution, binding events will occur at the Au-water interface, changing the SPP condition of the Au thin film. This change is then observed as a change in resonance angle. Conversely, when the molecule of interest detaches from the Au surfaces, it also changes SPP resonance condition. SPR sensors are often used to measure association and dissociation constants in addition to the simple binding assay in biological systems. This measurement can also be achieved by varying the incident wavelength to match the effective index of refraction. However, such a setup requires a high accuracy tunable laser source and so poses more technical challenges as compared to the simple angle measurement SPR systems.

The same SPR principle can be applied to localized surface plasmon resonances made available by confining the SPR to a small nanostructure. It is possible to monitor the change in the LSPR resonance condition using a very simple dark-field microscopy setup coupled to a simple desk-top spectrometer and achieve very high density parallel detection of molecular adsorption and desorption events.

Summary

With recent advancements in nanotechnology and materials science, SPP based devices are finding more applications. In addition to its broad applications in biomedical agents, sensors and enhanced spectroscopies, plasmonic devices can now be found in narcotic sensors, solar energy converters, field portable autoclaves and an expanding array of microelectronic devices. While most of these devices are still in early development stages, there is no doubt that the impact of SPP will expand as we gain better

understanding of its physics and a better handle on nanofabrication techniques.

References

1. Moskovits, M.: Surface enhanced spectroscopy. *Rev. Mod. Phys.* **57**, 42 (1985)
2. Atwater, H.: The promise of plasmonics. *Sci. Am.* **17**(13), 56–63 (2007)
3. Haslett, T.L., Tay, L., Moskovits, M.: Can surface-enhanced Raman scattering serve as a channel for strong optical pumping? *J. Chem. Phys.* **113**(4), 1641–1646 (2000)
4. Pillai, G.: Nanomedicines for cancer therapy: an update of FDA approved and those under various stages of development. *SOJ Pharm. Pharm. Sci.* **1**(2), 13 (2014)
5. O'Neal, D.P., Hirsch, L.R., Halas, N.J., Payne, J.D., West, J.L.: Photo-thermal tumor ablation in mice using near infrared-absorbing nanoparticles. *Cancer Lett. (Amsterdam, Neth.)* **209**(2), 171–176 (2004)
6. Fleischman, M., Hendra, P.J., McQuillan, A.J.: Raman spectra of pyridine adsorbed at a silver electrode. *Chem. Phys. Lett.* **26**(2), 4 (1974)
7. Jeanmaire, D.L., Duynne, R.P.V.: Surface raman spectroelectrochemistry: Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode. *J. Electroanal. Chem.* **84**(1), 1 (1977)
8. Moskovits, M.: Surface roughness and the enhanced intensity of Raman scattering by molecules adsorbed on metals. *J. Chem. Phys.* **69**(9), 4159 (1978)
9. Albrecht, M.G., Creighton, J.A.: Anomalously intense Raman spectra of pyridine at a silver electrode. *J. Am. Chem. Soc.* **99**(15), 5215 (1977)
10. Tay, L.-L., Hulse, J.: Surface-enhanced Raman and optical scattering in coupled plasmonic nanoclusters. *J. Mod. Opt.* **60**(14), 1107–1114 (2013)
11. Hirsch, L.R., Stafford, R.J., Bankson, J.A., Sershen, S.R., Rivera, B., Price, R.E., Hazle, J.D., Halas, N.J., West, J.L.: Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. *Proc. Natl. Acad. Sci. U. S. A.* **100**(23), 13549–13554 (2003)
12. Hu, Q.Y., Tay, L.L., Noestheden, M., Pezacki, J.P.: Mammalian cell surface imaging with nitrile-functionalized nanoprobe: biophysical characterization of aggregation and polarization anisotropy in SERS imaging. *J. Am. Chem. Soc.* **129**(1), 14–15 (2007)

T

Tactile Painting and Haute Couture

- [Art and Fashion Color Design](#)

Temporal Collective Color Preferences

- [Color Trends](#)

Tendencies

- [Color Trends](#)

Tessellations

- [Mosaics](#)

Tesserae

- [Mosaics](#)

Tetrachromatic Vision

Gabriele Jordan¹ and John D. Mollon²

¹Biosciences Institute, Newcastle University,
Newcastle upon Tyne, UK

²Department of Experimental Psychology,
Cambridge University, Cambridge, UK

Synonyms

[Four-dimensional color vision](#)

Definition

Tetrachromatic color vision here refers to human color vision that relies on the presence of four types of retinal cone photopigment whose signals are processed independently. By analogy with the classical, psychophysical, definition of human trichromacy [1], a tetrachromatic observer will require four primaries in a color-matching experiment to match any other color. In comparative studies of color vision, the term tetrachromacy usually refers to the presence of four distinct types of photoreceptor [2].

A distinction has been made between strong and weak tetrachromacy [3]. In both cases, there are four types of cone in the retina, but only the former denotes behavioral tetrachromacy, where the cortex gains independent access to a fourth signal allowing the individual to discriminate

colors along a dimension denied to color-normal people.

Conceptual Background

In the case of stimuli presented to the fovea, normal human color vision is trichromatic in that three variables or primary lights are needed to match any given color [1]. Since Thomas Young’s Bakerian Lecture of 1802, it has been thought that the basis of trichromacy is the presence in the retina of three types of “sensitive filament,” now known to be the cone photoreceptors. According to the wavelength of maximal sensitivity, these cones are referred to as S (short-wave sensitive), M (middle-wave sensitive), and L (long-wave sensitive).

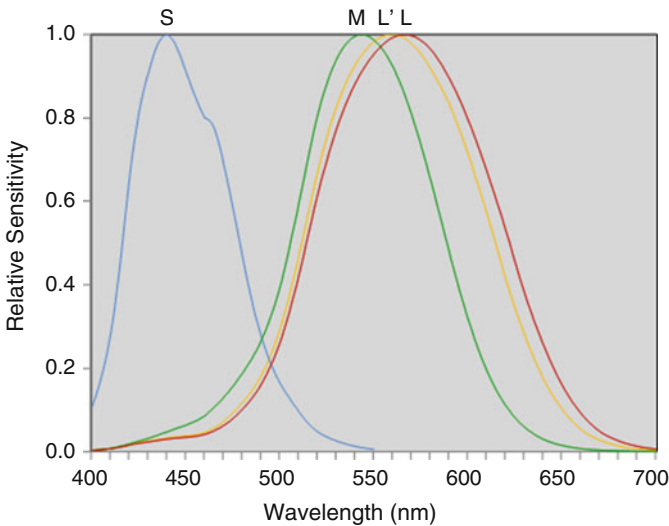
To explain the existence of additional types of cone with different spectral sensitivities (M’ or L’), one needs to consider the genetics of the M and L cone photopigments (for review see [4]). Nathans et al. [5] sequenced the M and L protein genes that are located in a tandem arrangement on the X chromosome and found a 98% homology of their DNA. This has been taken as evidence for an evolutionarily recent duplication of an ancestral gene, followed by some degree of divergence of the sequences of the two original copies. Highly similar genes tend to

misalign during meiosis and recombine in different ways. A common outcome of this process is a hybrid gene that consists of part L and part M gene. Importantly, such a hybrid gene will express a photopigment with a spectral sensitivity intermediate to those of the normal L and M cone photopigments [6].

Some 6% of men have such a hybrid gene on their single X chromosome and, as a consequence, will be classified as anomalous trichromats. Specifically, 5% of these men are deuteranomalous (Da), and their color vision relies on inputs from S, L, and L’ cones, while another 1% of men are protanomalous (Pa) having S, M, and M’ cones in their retinæ. Of interest here are the first-degree female relatives (mothers, daughters) of anomalous trichromats. These women are heterozygous for either deutan or protan anomaly and are candidates for tetrachromacy: On one of their X chromosomes, they have the genes for the normal L and M cone photopigments, but on the other they carry a hybrid gene that encodes a spectrally shifted cone photopigment. Figure 1 shows the four normalized spectral sensitivity curves for the S, M, L’, and L cones of a hypothetical carrier of deuteranomaly (cDa).

An early paper by Mollon and colleagues [7] provides a useful analogy to human tetrachromacy. They studied variations of color vision in a basically dichromatic species of New World

Tetrachromatic Vision,
Fig. 1 Normalized spectral sensitivity curves for the S, M, L’, and L cones of a hypothetical carrier of deuteranomaly (cDa)



primate, the squirrel monkey. They discovered in the population pool a range of cone pigments with different spectral sensitivities and, more importantly, behavioral trichromacy in a subset of females. Though no genetic analyses were available at the time, the authors hypothesized – correctly as it turned out – that the origin of these animals' good color discrimination was genetic and afforded to them by heterozygosity at the X-linked locus.

X-Chromosome Inactivation and the Retinal Cone Mosaic

In females, during embryonic development, one of their two X chromosomes is silenced [8]. This process is random with respect to the parental origin of the X chromosome, but once inactivation has taken place in a given cell, it is preserved in all daughter cells. Crucially for tetrachromacy, the different types of expressed photopigment will be segregated in different cone cells so that the retina of a carrier for, say, deuteranomaly will be a mosaic of normal and anomalous cones (S, M, L', and L), and the size of the retinal cone patches will depend on the time of onset of X inactivation and the migration of cones in the developing retina. The analogous case exists for carriers of protanomaly (cPa) though they are much rarer in the population.

What Are the Necessary Conditions for Tetrachromacy?

To allow tetrachromacy, it would seem that two essential conditions should be met: Firstly, in the middle- to long-wave region of the spectrum, the heterozygote must have three photopigments that are sufficiently different in their spectral absorption curves. Secondly, there must be independent neural channels that preserve the ratios of excitations of these photopigments.

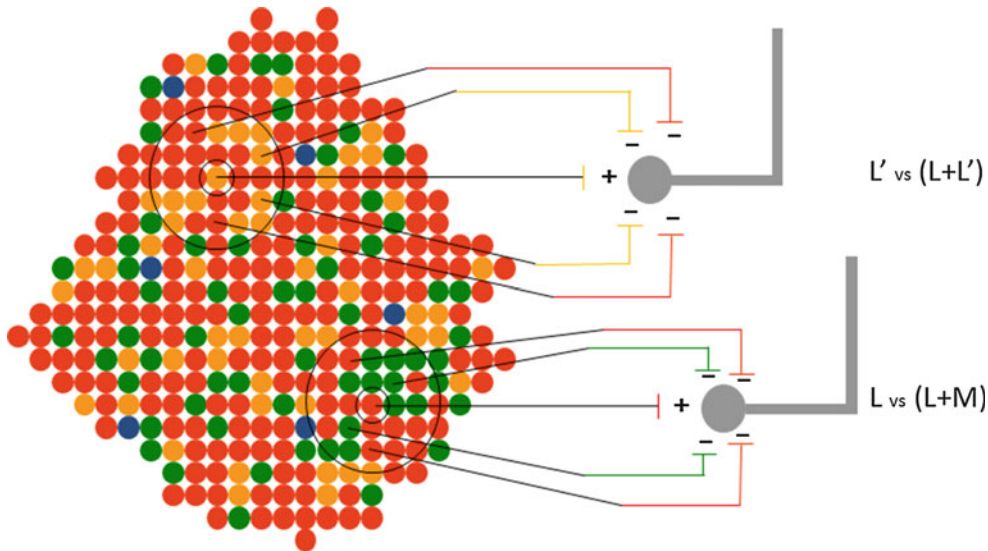
- (i) It is known that there is a range of L pigments with different peak sensitivities and similarly a range of M pigments [9]. In particular, a

common polymorphism at site 180 in the L cone photopigment gives an ~ 4 nm spectral difference in sensitivity [4, 5]. Jameson and colleagues have suggested that this polymorphism may indeed underlie tetrachromacy [10]. If true, tetrachromacy would then be predicted to occur in about 47% of the Caucasian population and the question arises as to whether such a relatively small spectral separation between two L photopigments would be sufficient to provide an additional, salient signal [11].

In addition, some hybrid pigments may be present in lower optical density and thus have narrower absorption curves [12]. How big a difference in absorption curve between, say, L and L' is sufficient for tetrachromatic color vision? It may be significant that the strongest candidate for tetrachromacy identified so far (see below [13]) has L and L' cone photopigments with an estimated spectral difference of 12 nm.

- (ii) Color vision does not simply depend on the number and type of retinal cone photopigment. In order to perceive colors and be able to discriminate between them, the inputs from different classes of cone need to be compared and two chromatic, post-receptoral channels were classically identified for this task (for discussion see [14]).

Of particular interest for tetrachromatic color vision is the phylogenetically younger, “red-green” opponent channel. Cells of the parvocellular pathway have receptive fields that are divided into spatially and chromatically antagonistic center and surround regions: In the fovea, midget ganglion cells draw their center input from a single cone of one type (e.g., L or M), whereas the surround input is likely to be drawn indiscriminately from L and M cones (mixed surround) (for review see [15]). Such a channel lends itself to tetrachromatic color vision as its chromatic specificity is determined primarily by the type of cone feeding into the center. Thus, if the signal from a single L' cone is compared to the pooled input of other cones, say $L + L'$, as is the case in Da, a cDa individual too will have the neural basis of making



Tetrachromatic Vision, Fig. 2 Schematic representation of a tetrachromatic cone mosaic with a “deuteranomalous” and a normal red-green channel. For explanation see text

similar comparisons. Figure 2 shows a schematic representation of a tetrachromatic cone mosaic with a deuteranomalous (L' vs $(L + L')$) as well as a normal (L vs $(L + M)$) red-green channel. If both channels are capable of providing salient input to cortical mechanisms, then strong tetrachromacy should result.

Historical Considerations

In the first half of the twentieth century, the majority of scientists concerned with individual differences in human color vision and possible phenotypic manifestations of X-linked heterozygosity believed that female carriers of red-green deficiencies either shared a little in their sons' disability or did not differ in any significant way from color-normal observers (for review see [3]). A different view was held by the Dutch physicist De Vries who, in 1948, explicitly stated that the daughters of a Da father “must be tetrachromatic,” and that it should be possible to demonstrate (strong) tetrachromacy using color mixing experiments. That the existence of an additional cone type in the retina indeed has an effect on color matches was found by Nagy and colleagues

33 years later by demonstrating failure of the additivity law [16]. According to this fundamental law of color matching for trichromats, the addition of a monochromatic light to an existing color match (here a match of a 546 nm and a 660 nm mixture ratio to a 588 nm reference light) should not upset the match even though the overall appearance might be different [1]. Four carriers of an unidentified type, but no color-normal male or female control, changed their matches after chromatic adaptation. Since the carriers were nevertheless capable of making trichromatic matches, they would be classified as weak tetrachromats.

Methodological Issues and Results for Strong Tetrachromacy

Explicit tests for strong tetrachromacy must ensure that a fourth chromatic primary is provided on which the tetrachromat can base her discriminations. Thus, no ordinary RGB color monitor can be used for any test of this kind.

Ideal is a specifically designed colorimeter that allows inclusion and manipulation of the wavelength of n primary lights with independent control of luminance. It would also permit the precise

specifications of spatial and temporal stimulus parameters. Such a system was used by Jordan et al. [13] for the identification of tetrachromats in a Rayleigh-type discrimination test. The stimuli used were in the spectral range from 556 to 670 nm, where S cones are insensitive. A triplet of three successive lights was presented. One light was a red-green mixture (546 + 670 nm) that could vary in the proportion of red, and the other two were a monochromatic yellow (590 nm). Observers were asked to identify the red-green mixture. Since all normal trichromats can make a dichromatic color match in this spectral range, they were expected to fail to detect the mixture light for some red-green mixture and some intensity of the monochromatic yellow. In contrast, a strong tetrachromat was predicted to always detect the mixture: A given light mixture could simulate the excitations that the monochromatic yellow produces in her L and M cones, and a different mixture could simulate the excitations in, say, her L and L' cones; but no mixture of red and green light could concurrently satisfy both these equations. Only one out of 18 cDa (known as cDa29) and none of 7 cPa participants performed according to the prediction. As expected, all control subjects failed to discriminate. Since cDa29 could not make a trichromatic color match with only two primaries available in the Rayleigh region, she was classified as strong tetrachromat.

Another approach to explicitly test for strong tetrachromacy is to ask candidates to discriminate surface colors. The challenge here is that one needs to find stimuli that are metameric (indistinguishable) for normal observers but look distinct to a tetrachromat. Such a test was designed to demonstrate that Da individuals are capable of making discriminations not achievable by normal observers [17] and cDa individuals ought to be able to make the same discriminations if their L vs (L + L') channel is salient enough. Participants were asked to rate how distinct pairs of surface colors (two sets of different blue-yellow acrylic mixtures) looked. Multidimensional scaling (MDS) was then used to reconstruct the subjective color space for each observer. Four out of

nine cDa observers were found to be able to make discriminations on this test, and their MDS solution correlated significantly with the theoretical L vs (L + L') channel. One of these four carriers was cDa29. Molecular genetic analyses of saliva samples taken from these observers confirmed the existence of a hybrid gene for an L' cone [13].

To provide the most conclusive evidence for strong tetrachromacy, a full set of color-matching functions is needed for selected candidates. This would allow direct derivation of the four-dimensional color space of four-cone individuals. Furthermore, the influence of factors such as cone ratios, cone optical density, and absorption by the ocular media must be considered in further tests.

Tetrachromacy in Parafoveal and Peripheral Vision

In extra-foveal human vision, five types of photoreceptor are known to be present – rods, three types of cone, and intrinsically photosensitive ganglion cells, which contain melanopsin. Trichromatic simultaneous matches are possible in extra-foveal vision but are unstable with changes in light level or in adaptive state [1], whereas a tetrachromatic match can be found that is stable [18]. These are the characteristics that would be expected for weak tetrachromacy as defined by Jordan and Mollon [3]. It is true that Bongard and his colleagues [18] found that parafoveal trichromatic matches failed under the special condition where one test field replaced the other, but Brindley [1] describes the failure as a “momentary lightening of the field” at the transition. This could still be considered a case of weak tetrachromacy, since the state of adaptation changes at the moment of transition. Whether, at high light levels, a melanopsin signal contributes to subjective appearance in the periphery [19] is currently an issue of active investigation.

Cross-References

► [Color Vision Testing](#)

References

1. Brindley, G.S.: Physiology of the Retina and the Visual Pathway, pp. 204–205. Edward Arnold, London (1970)
2. Jacobs, G.H.: Photopigments and the dimensionality of animal color vision. *Neurosci. Biobehav. Rev.* **86**, 108–130 (2018)
3. Jordan, G., Mollon, J.D.: A study of women heterozygous for colour deficiencies. *Vis. Res.* **33**, 1495–1508 (1993)
4. Deeb, S.S.: Genetics of variation in human color vision and the retinal cone mosaic. *Curr. Opin. Genet. Dev.* **16**, 301–307 (2006)
5. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B., Hogness, D.S.: Molecular genetics of inherited variation in human colour vision. *Science*. **232**, 203–210 (1986)
6. Merbs, S.L., Nathans, J.: Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science*. **258**, 464–466 (1992)
7. Mollon, J.D., Bowmaker, J.K., Jacobs, G.H.: Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proc. R. Soc. Lond. B.* **222**, 373–399 (1984)
8. Lyon, M.F.: X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev.* **47**, 1–35 (1972)
9. Asenjo, A.B., Rim, J., Oprian, D.D.: Molecular determinants of human red/green color discrimination. *Neuron*. **12**, 1131–1138 (1994)
10. Jameson, K.A., Highnote, S.M., Wasserman, L.M.: Richer color experience in observers with multiple photopigment opsin genes. *Psychon. Bull. Rev.* **8**(2), 244–261 (2001)
11. Jordan, G., Mollon, J.D.: Tetrachromacy: the mysterious case of extra-ordinary color vision. *Curr. Opin. Behav. Sci.* **30**, 130–134 (2019)
12. Thomas, P.B.M., Formankiewicz, M.A., Mollon, J.D.: The effect of photopigment optical density on the color vision of the anomalous trichromat. *Vis. Res.* **51**, 2224–2233 (2011)
13. Jordan, G., Deeb, S.S., Bosten, J.M., Mollon, J.D.: The dimensionality of color vision in carriers of anomalous trichromacy. *J. Vis.* **10**(8), 1–19 (2010)
14. Mollon, J.D.: “Tho’ she kneeled in that place where they grew...” the uses and origins of primate colour vision. *J. Exp. Biol.* **146**, 21–38 (1989)
15. Thoreson, W.B., Dacey, D.M.: Diverse cell types, circuits, and mechanisms for color vision in the vertebrate retina. *Physiol. Rev.* **99**, 1527–1573 (2019)
16. Nagy, A.L., MacLeod, D.I.A., Heyneman, N.E., Eisner, A.: Four cone pigments in women heterozygous for color deficiency. *J. Opt. Soc. Am.* **71**, 719–722 (1981)
17. Bosten, J.M., Robinson, J.D., Jordan, G., Mollon, J.D.: Multidimensional scaling reveals a color dimension unique to ‘color-deficient’ observers. *Curr. Biol.* **15**(23), R950–R951 (2006)
18. Bongard, M.M., Smirnov, M.S., Friedrich, L.: The four-dimensional colour space of the extra-foveal retinal area of the human eye. In: *Visual Problems of Colour*, volume 1, pp. 327–330. Her Majesty’s Stationery Office, London (1958)
19. Horiguchi, H., Winawer, J., Dougherty, R.F., Wandell, B.A.: Human trichromacy revisited. *Proc. Natl. Acad. Sci. U. S. A.* **110**(3), E260–E269 (2012)

Textile Processing

► [Finish, Textile](#)

Textile Wet Processing

► [Finish, Textile](#)

Texture Measurement, Modeling, and Computer Graphics

Eric Kirchner

Color Technology Department, AkzoNobel Paints and Coatings, Zuid-Holland, The Netherlands

Synonyms

[Coarseness](#); [Glints](#); [Graininess](#); [Sparkle](#)

Definition

Following a definition by ASTM, texture can be defined as the visible surface structure depending on the size and organization of small constituent parts of a material, typically surface structure of a woven fabric. Many other materials that are used in the manufacturing industry also show texture, often of very different types. Examples are various types of wood, satin, paper, foods, and injection-molded plastics. This entry focuses on effect coatings, since the characterization of their

texture is advanced as compared to the case for most other materials. Also, effect coatings such as metallic and pearlescent coatings are widely used in the automotive, cosmetics, and packaging industry.

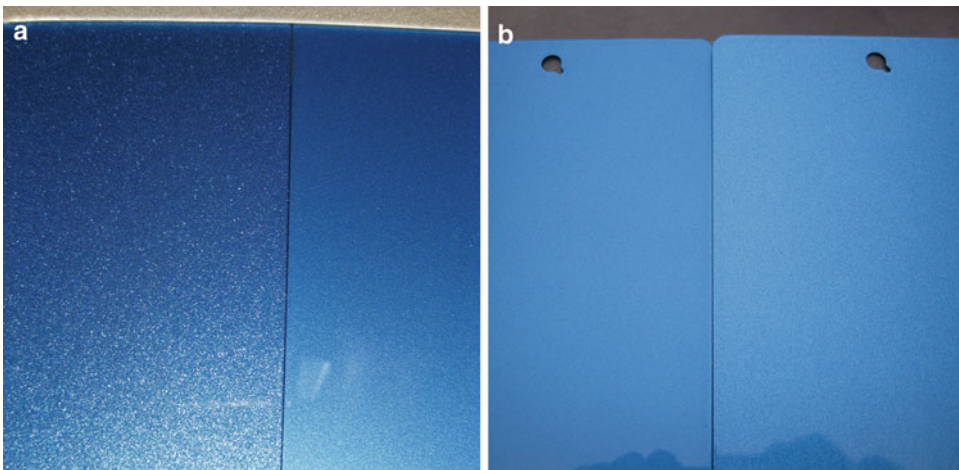
The texture of effect coatings is associated with the spatial variation in appearance caused by non-uniformity of colorant. This texture depends on the type of lighting, as shown in Fig. 1. Under intense unidirectional lighting, glint impression (also known as sparkle) is visible. This is the overall impression of several or many tiny light spots (“glints”) that are strikingly brighter than their surroundings. Under diffuse illumination conditions, the texture of effect coatings changes into diffuse coarseness (also known as graininess). This is the perceived contrast in the light/dark irregular pattern that is visible under diffuse illumination [1].

Overview

Conventional colorimetric results such as the modern color difference equations assume that the so-called reference conditions hold. The

Commission Internationale de l’Éclairage (CIE) recommended the investigation of non-reference conditions. According to one of the reference conditions, samples should be spatially uniform. Since effect coatings clearly do have a texture, their industrial importance explains the large body of research that has been conducted over the past decade on the texture of effect coatings.

Below, the progress is summarized that has been made on different subtopics in this area. Only recently, it became possible to objectively measure the texture of effect coatings. These developments are highlighted, since they also enabled the first quantitative studies on the influence of texture on perceived color differences between effect coatings. Because of these developments, it now has become possible to set tolerances on texture parameters. For color formulation and recipe prediction, models are needed that are able to predict texture values based on concentrations of colorants, similar to, e.g., the Kubelka-Munk theory for color prediction. Finally, it is discussed how the appearance of texture can be accounted for in computer graphics and 3D rendering of objects.



Texture Measurement, Modeling, and Computer Graphics, Fig. 1 (a) Two blue metallic paint samples differing in texture (sparkle, glint impression) as viewed under intense directional light. (b) Two blue metallic paint samples differing in texture (graininess, coarseness) as

viewed under diffuse light. Some other parameters for describing the texture of effect coatings have been introduced as well, and some of them will be discussed further below. However, those parameters generally lack strict definitions

Measurement and Observation of Texture

As long as no clear definitions of texture parameters were introduced and an understanding of the relation between illumination conditions and the type of texture of effect coatings was lacking, no commercial instrument for measuring texture could be developed. After having derived clear definitions and constraints on illumination conditions, a large amount of visual data was collected on different types of texture: diffuse coarseness for diffuse illumination conditions and glint impression for intense directional lighting. This body of data was obtained for hundreds of different effect coatings observed under different angles and illumination conditions. The visual data was correlated with texture parameters that were obtained from image analysis applied on CCD images of the same samples. In this way, algorithms were obtained to train and test prototypes of an instrument [1, 2].

The resulting instrument is the BYK-mac™, which is commercially available from BYK-Gardner since 2007. Similar studies are being conducted by other manufacturers of optical instruments as well [3], whereas also the development of the underlying image analysis algorithms has been taken up by the academic world [2].

Based on feedback from the automotive OEM industry, the set of texture parameters considered to be best applicable for the OEM industry has gradually evolved over the years. A graininess parameter G_r was introduced that has a high correlation with the diffuse coarseness parameter mentioned before. Apart from that, glint impression (sparkle) is considered to have two separate dimensions, called sparkle intensity S_i and sparkle area S_a [4]. An overall parameter sparkle grade (S_G) is no longer thought to be important.

Recently, also lighting booths became available that focus on observing the different types of texture under various geometries, such as the byko-spectra effect light booth.

In 2018, a second commercial instrument to measure texture including sparkle and coarseness,

independent from the BYK-mac™, was launched by X-Rite under the name MA-T12™ (a derived instrument with fewer measurement geometries for reflectance measurements goes by the name MA-T6™).

One major hurdle for further advancement in measuring texture is the absence of a widely accepted set of samples that can act as a scale for sparkle (and similarly also for coarseness/graininess). Such scales were developed as part of the research that led to the BYK-mac instrument [1, 5], but these scales did not become publicly available. In the past few years, a special committee of the CIE has defined a preliminary scale for sparkle and graininess, and the first research publications with these scales show promising correlations between measurement data and visual observations [6]. The CIE intends to conduct further research to determine if these scales need further fine-tuning. Part of this work also aims at developing image analysis techniques to extract numerical values for texture parameters [1, 2, 6], using more generic image processing algorithms for texture analysis.

Influence of Texture on Color Perception

Using artificially generated images of textures, several studies showed that the presence of texture has an influence on perceived color differences. Images exhibiting textured lightness variations have been shown to increase the visual tolerance for lightness deviations. This explains why, for effect coatings, a color difference equation like dE_{CMC} is best used with $l = 1.5$; $c = 1.0$; for solid (untextured) colors, best results are typically obtained for $l = 1.5$; $c = 1.0$.

However, quantitative studies on effect coatings only became possible after 2007 as outlined in the previous section. Based on a large study on 405 effect sample pairs, an expression was derived that quantitatively describes how differences in texture and color combine into a total appearance difference. The validity of this approach was confirmed in later research [5].

Setting Tolerances on Texture

Being able for the first time to measure texture parameters of effect coatings, manufacturing industries decided to include consistency in texture in their criteria for quality control. Among the first industries to investigate how to set industrial pass-fail tolerances on texture are the automotive industry and the plastics industry, where texture matching used to be a long-standing problem for quality control.

Most of the global automotive manufacturers are currently setting texture tolerances, as evidenced, for example, by multiparty meetings of the Detroit Color Council in the USA and BYK-Gardner user meetings in Europe. However, no specific tolerances on texture parameters have been published yet.

Based on perception thresholds of texture parameters, a proposal for realistic tolerances was recently published for different sets of texture parameters: diffuse coarseness, glint impression, graininess, and sparkle grade [7]. Only for the parameters sparkle area and sparkle intensity, the lack of published data makes it not yet possible to derive tolerance values.

Predictive Models for Texture

For color matching, optical models are widely used to optimize color formulations. In the 1940s, Kubelka and Munk developed a model that is still widely used in the paint and textile industry. In order to optimize the concentrations of colorants such that also the texture of an effect coating is matched, a similar model for texture is needed. Studies show that texture parameters like sparkle intensity and graininess can vary greatly with paint composition [8]. However, an accurate optical model for predicting texture based on a physical model like Kubelka-Munk has not been developed yet. The first example of an accurate model for texture uses a very different derivation [9]. It is based on a statistical approach for finding those mathematical terms that contribute

significantly to the texture properties that are to be predicted.

A mathematical-physical model to predict the visibility of sparkle as a function of viewing distance, metallic flake size, and other parameters has been recently developed in analogy to astronomical publications related to the visibility of stars during twilight [10].

Rendering of Texture in Computer Graphics

Digital rendering of three-dimensional objects has become very common in the cinematic and computer game industries. However, for accurate rendering of effect coatings also, the texture aspect needs to be taken into account. Most current approaches either completely ignore texture aspects or utilize a very crude and nonquantitative approach by adding sparkles to the image afterward [11].

Recently, two different approaches were published that do make accurate texture rendering possible. Using a few measured texture values as input as well as reflection measurements from an ordinary multiangle spectrophotometer, it was shown that accurate renderings can be produced [12]. No full BRDF and BTf measurements are needed. The core algorithm was derived from procedures used for determining texture parameter values based on images captured from effect coatings. By reversing these procedures, an algorithm was obtained that produces images resembling effect coatings, based on measured texture parameter values. The algorithm does not aim at simulating the reflection from each and every flake in a scene, but it accounts for their combined effect in an effective way.

In contrast to conventional methods for rendering, the derivation of the algorithms and the verification of the accuracy of the visualizations were both determined using visual tests in which rendered images were directly compared to physical objects. Thus, a truly appearance-based method for rendering color and texture was obtained, with

greatly reduced demands on computation power and disk storage.

In the past few years, a large number of publications have appeared related to the accurate visualization of sparkle, originating from groups working at the Universities of Berkeley and Zürich [13], Alicante, CSIC Madrid, and AkzoNobel [12], the Czech Academy of Sciences and Schlenk [14], and others.

Future Directions

Despite all the work that has been done on texture research over the past decade, this area is still in development. There is a clear need for more accurate image analysis algorithms to extract parameters from CCD images that correlate well with visually perceived texture. For manufacturers of optical instruments, this is an absolute necessity for developing instruments that can form a next generation after the commercial launch of the BYK-macTM in 2007.

Perception studies on the effect of texture on perceived color, or on perceived color differences, is also expected to be an active area of research for the coming decade. Currently, color difference equations are used that are derived from very different types of samples, such as solid colors or textiles. At present, texture can only be accounted for through parametric factors.

Setting tolerances on texture parameters and predicting texture properties based on paint composition are both areas of great industrial importance. Yet, these areas have hardly been explored yet.

For accurate rendering of textured paints, there are a few first publications, but also here the potential for improvement is enormous, and there are many obvious industrial applications in, e.g., the movie and the game maker industry.

Finally, it is good to mention that other types of textures appear in other types of paints. One example are powder coatings, where coarse and fine surface textures are common also in coatings without effect pigments. Recently the first

instrument to digitally measure and control also these textures has been commercially launched [15].

References

1. Kirchner, E.J.J., van den Kieboom, G.J., Njo, S.L., Supèr, R., Gottenbos, R.: The appearance of metallic and pearlescent materials. *Col. Res. Appl.* **32**, 256–266 (2007)
2. Kitaguchi, S., Westland, S., Kirchner, E.J.J., van den Kieboom, G.J., Luo, M.R.: Computational model for perceptual coarseness prediction. In: *Proceedings Conference on Graphics, Imaging and Vision*, pp. 267–282 (2006)
3. Ellens, M.S., Lamy, F.: From color to appearance in the real world. In: *Proceeding of Human Vision and Electronic Imaging XVII*, vol. 8291. SPIE, Burlingame (2012)
4. Klein, G.A.: *Industrial Color Physics*. Springer, New York (2010).; Kigle-Böckler, G.: New and innovative testing technologies for effect finishes. In: *Proceeding Conference of AIC*, Sydney (2009)
5. Huang, Z., Xu, H., Luo, M.R., Cui, G., Feng, H.: Assessing total differences for effective samples having variations in color, coarseness and glint. *Chin. Opt. Lett.* **8**, 717–720 (2010)
6. Ferrero, A., Perales, E., Basic, N., Pastuschek, M., Porrovecchio, G., Schirmacher, A., Velázquez, J.L., Campos, J., Martínez-Verdú, F.M., Smid, M., Linduska, P., Dauser, T., Blattner, P.: Preliminary measurement scales for sparkle and graininess. *Opt. Express.* **29**, 7589–7600 (2021). <https://doi.org/10.1364/OE.411953>
7. Kirchner, E.J.J., Ravi, J.: Setting tolerances on color and texture for automotive coatings. *Col. Res. Appl.* **39**, 88–98 (2014)
8. Rentschler, T.: Measuring sparkling blues without blues. *Eur. Coat. J.* **11**, 78–83 (2011)
9. Kirchner, E.J.J., Ravi, J.: Predicting and measuring the perceived texture of car paints. In: *Proceeding 3rd International Conference on Appearance. “Predicting Perceptions”*, Edinburgh (UK) (2012)
10. Kirchner, E., van der Lans, I., Perales, E., Martínez-Verdú, F., Campos, J., Ferrero, A.: Visibility of sparkle in metallic paints. *J. Opt. Soc. Am. A.* **32**, 921–927 (2015)
11. Meyer, G., Shimizu, C., Eggly, A., Fischer, D., King, J., Rodrigues, A.: Computer aided design of automotive finishes. In: *Proceeding of AIC Colour*, pp. 685–688 (2005)
12. Kirchner, E., van der Lans, I., Koeckhoven, P., Huraihat, K., Martínez-Verdú, F.M., Perales, E., Ferrero, A., Campos, J.: Real-time accurate rendering of color and texture of car coatings. *Proc. IS&T Int.*

- Symp. Electron. Imaging. **2019**, 076.1–076.5 (2019). <https://doi.org/10.2352/ISSN.2470-1173.2019.14.COLOR-076>
13. Yan, L.Q., Hasan, M., Jakob, W., Lawrence, J., Marschner, S., Ramamoorthi, R.: Rendering glints on high-resolution normal-mapped specular surfaces. *ACM Trans. Graph.* **33** (2014). <https://doi.org/10.1145/2601097.2601155>
 14. Filip, J., Vávra, R., Maile, F.J., Kolafova, M.: Framework for capturing and editing of anisotropic effect coatings. *Comput. Graph. Forum.* **40**, 68–80 (2021). <https://doi.org/10.1111/cgf.14119>
 15. Gibson, S.: AkzoNobel and BYK-Gardner digitize surface texture for powder coatings. *Polymers Paint Colour Journal* (April 2020). <https://polymerspaintcolourjournal.com/news/akzonobel-and-byk-gardner-digitise-surface-texture-for-powder-coatings>. Accessed 7 July 2021

The International Textile Machinery Exhibition (ITMA)

► [Coloration, Ink-Jet Printing](#)

Thermal Ink-jet (TIJ)

► [Coloration, Ink-Jet Printing](#)

Tincture

Paul Green-Armytage
The School of Media, Creative Arts and Social Enquiry, Curtin University, Perth, WA, Australia

Definition

The word “tincture” has more than one meaning in English. *The Concise Oxford Dictionary* has the following: “**1** a slight flavor or trace. **2** a tinge (of a color). **3** a medicinal solution (of a drug) in alcohol (*tincture of quinine*). **4** *Heraldry* an inclusive term for the metals, colors, and furs used in coats of arms.”

The word is used by specialists in particular fields and is uncommon in general conversation. However, some of the concepts associated with the word, if not the word itself, have potential value for those who are concerned with color and ► [appearance](#). In meaning 2, the word refers to very subtle colors. A reporter might describe a bride’s dress as having a beautiful tincture, white with a tinge of pink.

The word has a particular meaning in heraldry (meaning 4), and this draws attention to the lack of a word in general English usage that could embrace several different aspects of appearance in a single word. In heraldry, tincture includes color, luster, and texture. In coats of arms, the designs can be rendered in colors, “metals” (luster), and “furs” (texture). If tincture could be understood to mean all three aspects of appearance, the tincture of the bride’s dress could mean that the dress fabric also had a smooth texture and a lustrous sheen as well as being white with a tinge of pink.

In heraldry, the names used to identify the tinctures are Norman French. There are five “colors”: *gules* (red), *azure* (blue), *sable* (black), *vert* (green), and *purpure* (purple). Three other colors are found in rare cases and are sometimes referred to as “stains”: *sanguine* (blood red), *tenné* (tawny orange), and *murrey* (mulberry). The “metals” are *or* (gold) and *argent* (silver) – often represented by yellow and white. The “furs” are variations on the theme of ermine and vair (squirrel skins) and are represented by stylized patterns.

Where a figure, such as a cross, is set over a background, the “rule of tincture” is applied: the figure must not be from the same group of tinctures as the background. So a red (color) cross can be placed on a gold (metal) background but not on blue (color). This is to ensure maximum light–dark contrast and, therefore, maximum legibility and ease of identification.

There are lessons in heraldry for graphic designers today. The principles embodied in the rule of tincture are as relevant today as they were in the Middle Ages, and the concept of heraldic tincture is a reminder that there are other aspects

of ► [appearance](#), besides color, that can be exploited in design.

It has been suggested that the word “tincture” might be adopted and promoted with something like its heraldic meaning. “Tincture” could mean the totality of appearance characteristics, not only color, texture, and luster, as in heraldry, but also those other aspects of appearance that have been collected under the term “cesia.” “Cesia” is a term coined by César Jannello. It refers to ► [transparency](#), translucency, opacity, gloss, and specular reflection, each of which can also be represented in scales as from matte via semigloss to full gloss. And if the word “tincture” itself is problematic, with its meanings already established, perhaps another new word could be coined that would embrace all aspects of appearance: color, texture, and cesia.

Cross-References

- [Appearance](#)
- [Transparency](#)

Tint

- [Pigment, Ceramic](#)

Total Appearance

- [Appearance](#)

Traditional Color Model

- [RYB Color](#)

Traditional Color Theory

Zena O'Connor

Design Research Associates, Sydney, NSW,
Australia

Synonyms

[Color theory](#); [RYB color theory](#)

Definition

Traditional color theory represents a fit-for-purpose amalgam of grounded practical theory, normative design theory, and constructivism. It is underpinned by practical rather than scientific concerns and features a range of constructs, principles, and linked axioms that are largely programmatic. This branch of color theory emerged initially in respect to art practice, and as the professional practices of applied design and design of the built environment developed during the nineteenth and twentieth centuries, traditional color theory became a component of educational curriculum in these fields. However, during the mid- to late twentieth century, traditional color theory was rarely taught as a stand-alone subject but in tandem with practical design projects wherein students explored color application within a context that allowed individual creativity. Traditional color theory features the RYB color model, a conceptual model representing color creation using substance color and primary colors of red, yellow, and blue. A range of color constructs feature in traditional color theory including hierarchical color constructs and color combination constructs, relating to color application in art, applied design, and design of the built environment. Color-related constructs arising from traditional color theory represent a vocabulary frequently used in identifying and applying color in art, applied design, design of the built environment, and across diverse industry sectors. As such, these constructs serve as a *lingua franca*

for architects, artists, design professionals, and those involved with color application and within the broader community.

Introduction

A range of different color theories exist across diverse fields of research and inquiry including art, design, linguistics, phenomenology, physics, psychology, and vision science [1–4]. One of these is traditional color theory, which dates back some centuries and emerged initially in the field of art but gained traction in the fields of applied design and design of the built environment during the late nineteenth and twentieth centuries. As with all branches of theory, traditional color theory is characterized by various ontological and epistemological assumptions, theory-specific constructs, and traditions, plus it is characterized by a key focus on substance color. An evolving branch of color theory, constructs that emerged within traditional color theory in the past have come to represent a vocabulary frequently used among artists, designers, architects, and professionals working in color application across diverse industry sectors.

Defining Traditional Color Theory

The characteristics that distinguish historical traditional color theory and set it aside from other color theories include the following: (a) an ontological focus on substance color exploration and application in art, design, and architecture; (b) consideration of an exemplar set of primary pigment colors from which a large gamut of color nuances can be created; (c) a prescriptive approach to creating and applying color harmony; and (d) the provision of color combination guidelines for color exploration and application which are often used as color strategies in art, design, and architecture [3–6].

From a broad perspective, Green-Armytage identified distinct color categories relevant to color theory, and each of these tends to be underpinned by specific and diverse ontological and epistemological assumptions and traditions

[7]. These categories include conventional color, substance color, formula color, spectral profile color, psychophysical color, inherent color, and perceived color. The ontological focus of traditional color theory is primarily substance color in terms of pigments and dyes and, to a limited extent and by extrapolation, formula color.

The epistemological traditions of traditional color theory represent an amalgam of grounded practical theory, normative design theory, and constructivism [5, 6]. It is underpinned by practical rather than scientific concerns and features a range of constructs, concepts, and axioms that specifically relate to color identification, exploration, and application, as follows:

- a) Color is considered to have three attributes: hue, value, and saturation wherein.
 - a. Hue is the general color family cluster commonly referred to by general color terms such as “red,” “blue,” and “green.”
 - b. Value (also referred to as tone and tonal value) relates to the lightness or darkness (whiteness or blackness) of substance color.
 - c. Saturation (also referred to as chroma and chromaticity), relating to the chromatic intensity of color as represented by a continuum from full chromatic intensity to achromatic.
- b) Hierarchical color constructs include primary colors, secondary colors, and tertiary color.
- c) Color constructs that relate to color relationships and color application include analogous colors, complementary colors, monotone, monochromatic colors, and color chords.
- d) Conceptual color models that feature the RYB color triad and often represented by the two-dimensional RYB color model or three-dimensional versions.
- e) Guidelines for color exploration, color combination, and color application in art, design, architecture, and across diverse industry sectors.

Traditional color theory has continued to evolve over time as evidenced by a shift in focus from a prescriptive approach to color combination for application in art, design, and architecture (e.g., the color theories of Munsell) to an

exploratory approach, as per the color theories of Itten and Albers. This shift represented rebuttal of a deterministic, nomothetic understanding about color (relating to color combination, color application plus responses to color) in favor of a post-positivist, idiographic understanding about color, color combination, and application.

As a result of this shift, prescriptive color combination and color harmony principles and formulae were generally replaced by an approach that acknowledges the complexity of color and which feature a stronger focus on color exploration. For example, both Itten and Albers advocate color exploration, with Albers' suggesting exploratory "practice *before* theory," noting that prescriptive color guidelines and formulae are, "...worn out. No mechanical color system is flexible enough to pre-calculate the manifold changing factors ... in a single prescribed recipe." [8–10] This shift also acknowledged Hård and Sivik's assertion that the number of possible color combinations "is almost infinite," highlighting the futile nature of prescriptive color combination guidelines and formulae for color application [11].

Origins and Development of Traditional Color Theory

Aside from commentary offered by ancient Greek philosophers, Tobias Mayer is credited with conceptualizing and documenting the first RYB color model that subsequently served as a basis for development of traditional color theory [2, 4]. Multiple traditional color theorists since Mayer emerged over time, and these generally featured the RYB triad supported by rationale and discussion. These color theorists include Moses Harris, Philipp Otto Runge, Mary Gartside, George Field, Charles Hayter, J.L.F. MÉRIMÉE, Michel-Eugène Chevreul, Albert Munsell, Johannes Itten, and Josef Albers.

As the nineteenth century segued into the twentieth century, the ontological focus of these color theories remained substance (pigment color), including specific types of pigment color such as gouache paint. In this context, gouache (and earlier variations including body color and

opaque watercolor) had the capacity to create a large gamut of color nuances via RYB triad intermixture and create large, flat areas of opaque color with "hard" edges suitable for professional design output. Research for this entry indicates that numerous works of prominent designers and architects at the time used gouache paint including designs and renderings by Owen Jones, Charles Rennie Macintosh, Henry Van Der Velde, William Morris, Christopher Dresser, and Frank Lloyd Wright plus the majority of lecturers and students at the Bauhaus and subsequent design and architecture schools.

The preliminary course at the Bauhaus, which was undertaken by all students, was initially developed by Johannes Itten. In this course, students explored theories of color, design, and space within the context of practical application exercises. Underpinned by an experiential approach, the curriculum allowed for individual creative expression in tandem with subjective and objective observation. As such, it was acknowledged at the Bauhaus that color did not exist in a vacuum, and hence, traditional color theory was not taught as a stand-alone subject from a purely theoretical perspective. Instead, students were encouraged to explore color relationships and constructs common in traditional color theory within the context of practical color design applications relevant to applied design and architecture [12]. Anecdotal research suggests that the pedagogical approach of the Bauhaus, wherein traditional color theory was taught in context of practical color design application, became the norm, and various iterations of this approach featured in the curriculum of applied design and architecture schools across the Western world including Parsons School of Design (New York) and the Shillito Design School (Sydney). However, aside from the publications of Itten and Albers, plus various textbooks, specific color curriculum details from twentieth century remain largely undocumented to date.

Despite its practical focus, traditional color theory did not feature a structured and documented methodology to identify and communicate color information across different applications or industry sectors. However, a number of

theorists developed practical color atlases based on RYB color including Johann Heinrich Lambert in the eighteenth century, who had the foresight to anticipate the benefits of his *Farbenpyramide* (Color pyramid, 1772) for artists, merchants, and design applications [4]. Similarly, David Ramsay Hay published *A Nomenclature of Colours* (1845) which featured 240 colors based on RYB color [2, 4]. Many of the early color atlases were not widely available and had a limited focus, such as entomology or ornithology. In the mid-twentieth century, Jean-Philippe Lenclos developed a methodology to collect and identify color data relating to the built environment, a methodology that prompted the development of the environmental color-mapping technique. Color matching in art, design, and architecture was conducted manually, and the unreliability of this process proved problematic for those applying color across diverse applications due to color mismatch issues. As a result, color notation systems were developed to aid identification and specification of color in industry including the Munsell color notation and color atlas (initially published 1905, 1913), the RAL color system (launched 1927), the Pantone PMS system (launched 1963), and the NCS color system (launched 1979) [5, 13, 14].

Hierarchical Color Classifications

Traditional color theory and the RYB color model are underpinned by specific hierarchical color classifications. The first level is referred to in the literature as “primary,” “principal,” “simple,” or “basic” colors and represented in the RYB color model as red, yellow, and blue. The next level is referred to as “secondary” or “compound” colors, representing intermixture of these primary colors, and identified in the RYB color model as orange, green, and purple. A third hierarchical level is referred to as “tertiary” color, representing an intermixture of red, yellow, and blue. From the twentieth century onward, these intermixtures do not necessarily represent strictly proportional synthesis of RYB pigment colors; instead hierarchical color classifications represent opportunities to

explore color nuances rather than prescriptive formulae for color creation [2, 3, 5].

In this context, primary, secondary, and tertiary colors represent exemplar hues as opposed to actual hues. In respect to the hierarchical color classification of “tertiary” color, Ruskin muddled the common understanding of the term by suggesting that “tertiary” colors were those that lay *between* secondary colors as this enabled him to avoid what he termed *tiresome* color names [5, 15]. This subordinate understanding of the term “tertiary” color has continued due to Ruskin’s solecism.

RYB Color Model and Illustrations

One of the earliest RYB color illustrations were the hue circles of Moses Harris, published in *The Natural System of Colours* (Fig. 1).

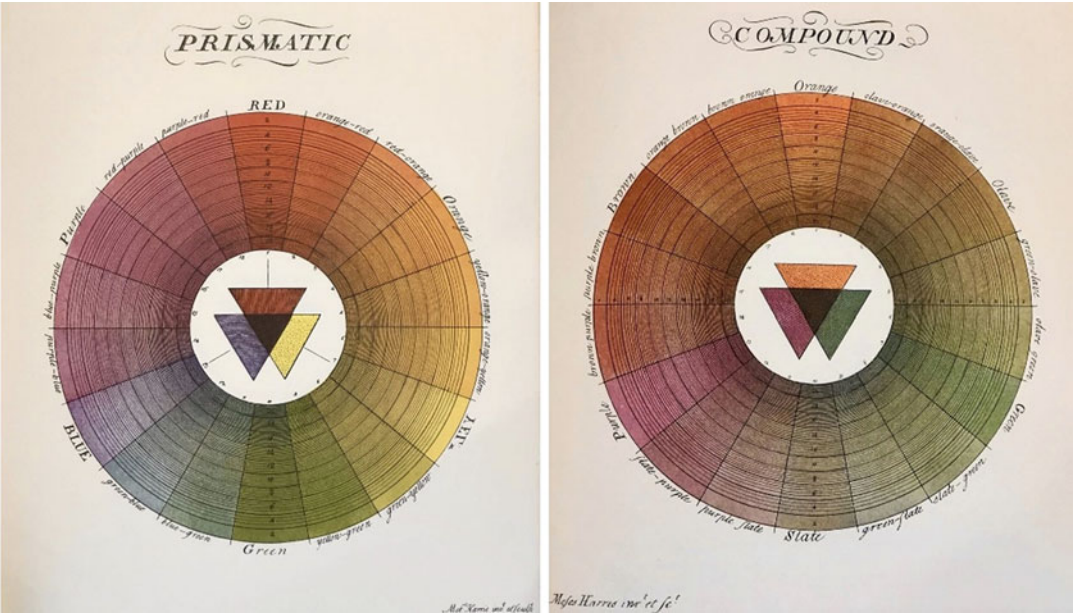
Subsequent traditional color theorists developed similar hue circles and color wheel models including Charles Hayter and George Field. Perhaps the most recognizable color models were those of Albers and Itten, and it is primarily Itten’s color wheel and color star model that often featured as teaching and learning tools in color curriculum that focused on traditional color theory in art, applied design, and architecture schools (Fig. 2).

The RYB color model is a key conceptual model that underpins traditional color theory. It is not intended to illustrate the complexity of color or the multitude of hue nuances achievable via pigment color intermixture but an illustrative model of basic color creation and intermixture.

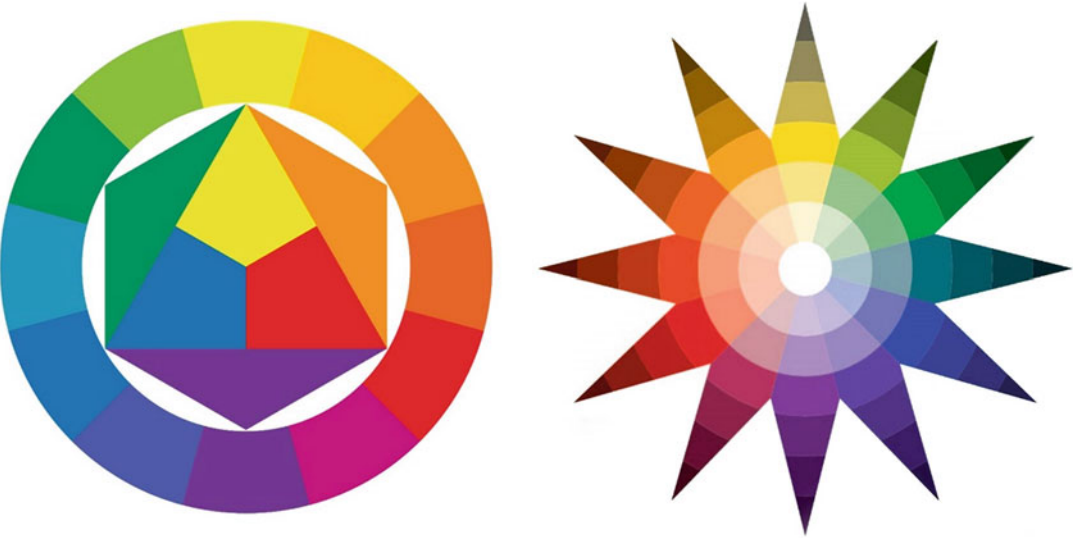
The RYB color model reflects a similar format of other color models such as RGB and CMY. However, each of these color models is distinctly different in terms of ontology and epistemological traditions and should not be conflated or used substitutionally in any way.

Traditional Color Theory Constructs

A distinguishing factor of nineteenth- and twentieth-century traditional color theories are various constructs intended for color exploration



Traditional Color Theory, Fig. 1 Moses Harris hue circles, 1766 (Illustration facsimile edition, 1963)



Traditional Color Theory, Fig. 2 Itten color wheel model and color star, 1961 (Facsimile illustrations by Zena O'Connor and Geoffrey Gifford, 2021)

and application. These constructs, which tend to differ slightly in definition, often represented guidelines for color exploration in an educational context.

Multiple color constructs, which featured in color curriculum and related textbooks, continue to serve as a lingua franca among artists, designers, architects, and professionals involved

with color application across diverse industry sectors [5, 6].

Aside from the hierarchical color classifications of primary color, secondary color, and tertiary color, additional color constructs that relate to hue include but are not limited to the following:

- Analogues colors – these are colors that share a degree of hue similarity but may differ in tonal value and saturation levels. In early traditional color theories, this construct was referred to as “harmonious” colors, a term that has largely been discarded.
- Complementary colors – also referred to as contrasting colors, these are hues that generally sit on opposite sides of color wheel models within the context of traditional color theory.
- Monochromatic – refers to one hue illustrated in a range of tonal value and saturation levels.
- “Warm” colors and “cool” colors – representing discrete groups of colors that are associated with the appearance of warm or cool contexts, substances, illustrations, or themes.
- Tertiary color wheel which may or may not feature three primary colors at the center surrounded by secondary colors, plus an array of tertiary colors created via intermixture of secondary colors.

Color constructs that relate to tonal value and are often used in color exploration and application include the following:

- Monotone – refers to a range of hues that share similarity of tonal value.
- Shadow palettes – refer to combining groups of colors that share similar tonal values and which have been reduced in hue via the addition of gray.
- Tints and shades – refer to variations of tonal value relative to a set of hues. Tints involve the addition of white, and shades refer to the addition of gray, black, or a contrasting color.
- Key – this term and “chord” have been appropriated from music theory. It refers to the quality of tonal value inherent in a color from *high key* referring to light tonal values, *intermediate key* relating to midtonal values, and *low key* referring to dark tonal values.

- Chords – relate to intervals between the tonal values evident in a group of colors. A broad range of tonal values is referred to as *major chord* and a narrow range as *minor chord*. This terminology is common among those specifying and applying color, including typical examples like *high key*, *major chord* and *low key*, and *minor chord*. In visual communications design and interior design, various chord and key combinations are considered to convey different impressions in terms of ambience and mood.
- Natural tonal value order – the notion that colors that frequently feature on color wheel models exhibit a “natural” tonal value, that is, their tonal values tend to be common in nature and, hence, have a relatively high level of familiarity. When reversing these tonal values from high to low, or vice versa, they are considered to appear unfamiliar, “unnatural,” and therefore discordant.

Additional color constructs that occur in traditional color theories often represented by color effects and included in traditional color theory curriculum as a way of exploring color complexity include the following:

- Discords – generally defined as groups of two or more colors that feature tonal values wherein their tonal value has been reversed according to the “natural tonal value order”.
- Broken color – generally defined as a relatively pure pigment color that has been “broken” or diluted by the addition of white, black, gray, or contrasting color, or alternatively, the overlay of colored texture.
- Dirty color – creating and observing the effect that contextual color may have in reducing the apparent purity of pigment color.

Illustrations that were a relatively frequent occurrence in traditional color theory and which were often the basis for guided color exploration in an educational context include the following:

- Color wheel models and color star models similar to those of Itten.

- Value scales representing achromatic grays at equidistant intervals between white and black. These scales often featured nine-step intervals that included white and black.
- Tonal value wheel which features hues drawn from a color wheel model plus corresponding grays that feature the same tonal value of each individual hue.
- Tonal value charts of a range of different individual hues.
- Tints and shades chart which generally feature a range of hues plus light tonal values and dark tonal values of each hue.
- Analysis of a painting – this exercise devised and implemented by Albers and Itten involved identifying the colors evident in a painting in terms of hue, tonal value, and saturation.

Components of traditional color theory continue to feature in curriculum at art, design, and architecture schools. However, the emergence of digital color technologies has altered the way in which color is created and applied within educational contexts and across diverse industry sectors. Despite this, constructs that emerged in traditional color theory continue to be commonly used and frequently applied by artists, designers, architects, and color design professionals [5, 6].

Cross-References

- [Chevreul, Michel-Eugène](#)
- [Color Combination](#)
- [Color Harmony](#)
- [Complementary Colors](#)
- [Itten, Johannes](#)
- [Mayer, Tobias](#)
- [Munsell, Albert Henry](#)
- [Primary Colors](#)
- [Runge, Philipp Otto](#)
- [RYB Color](#)

References

1. Gage J.: Colour at the Bauhaus. AA Files, No. 2, July, 50–54 (1982)

2. Gage, J.: Colour and Culture. Thames & Hudson, London (1995)
3. Gage, J.: Colour and Meaning: Art, Science and Symbolism. Thames & Hudson, London (2000)
4. Kuehni, R.G., Schwarz, A.: Color Ordered: A Survey of Color Order Systems from Antiquity to the Present. Oxford University Press, Oxford (2008)
5. O'Connor, Z.: Traditional colour theory: A review. Color. Res. Appl. **46**, 838–847 (2021)
6. O'Connor, Z.: A defence of traditional colour theory: Exploring the impact and relevance of ontology and epistemology. In: Color Impact 2020 Symposium, Point/Counter Point online debate, 6 June 2020
7. Green-Armytage, P.: The value of knowledge for colour design. Color. Res. Appl. **31**(4), 253–269 (2006)
8. Itten, J.: The Art of Color (Revised Edition, 1973). Wiley, New York (1961)
9. Itten, J.: Design and Form: The Basic Course at the Bauhaus and Later. Wiley, New York (1975)
10. Albers, J.: Interaction of Color (50th Anniversary Edition). Yale University Press, New Haven (2006)
11. Hård, A., Sivik, L.: A theory of colors in combination – A descriptive model related to the NCS color-order system. Color. Res. Appl. **26**(1), 4–28 (2001)
12. Bayer, H., Gropius, W., Gropius, I. (eds.): Bauhaus: 1919–1928. Secker & Warburg, London (1975)
13. Pantone: About Pantone, 2020. <https://www.pantone.com/about/about-pantone>. Accessed 6 May 2020
14. NCS: NCS – our history, 2020. <https://ncscolor.com/about/our-history/>. Accessed 6 May 2020
15. Ruskin, J.: The laws of Fesole: A familiar treatise on the elementary principles and practice of drawing and painting; as determined by the Tuscan Masters; Arranged for the Use of Schools, vol. 1. George Allen, Kent (1879). <https://archive.org/details/lawsfssoleafamil03ruskgoog/>. Accessed 12 Aug 2020

Transformation of Color Matching Functions

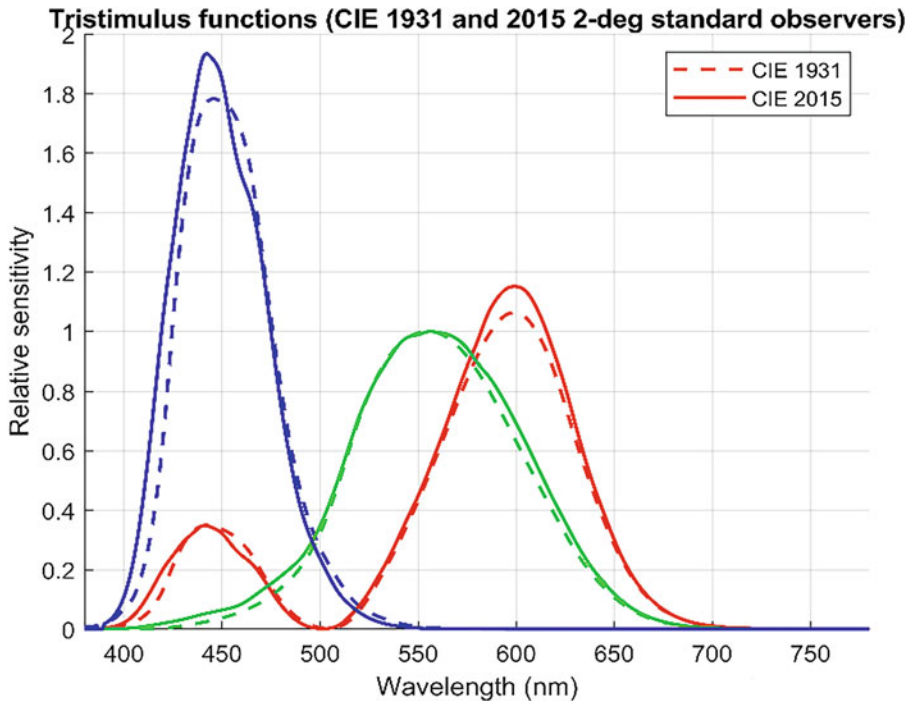
Lindsay W. MacDonald
Faculty of Engineering Sciences, University
College London, London, UK

Synonyms

[Computation of colorimetric stimuli](#)

Definition

Conversion of a colorimetric stimulus from one observer to another.



Transformation of Color Matching Functions, Fig. 1 Comparison of the tristimulus functions of the standard observer defined by the CIE 1931 and CIE 2015 systems of colorimetry

Introduction

Because the X, Y, Z tristimulus functions in the CIE 1931 and 2015 specifications [1, 2] have been derived from different representations of the cone response functions, they do not have a linear relationship (see Fig. 1). In particular, the CIE 1931 functions underestimate the sensitivity for wavelengths below 460 nm. Thus, it is not possible to express an exact relationship between the two sets of tristimulus functions, nor between any corresponding triplets of tristimulus values, by means of a system of equations. Instead, it is necessary to use linear regression to determine an approximation between large numbers of corresponding X, Y, Z values derived from the two sets of basis functions.

Fitting with Real Reflectance Spectra

A dataset of 8714 reflectance spectra was collated from readily available spectral reflectance

measurement sets of real material surfaces over the wavelength range 380–780 nm, with measurement data at intervals of 10 nm or less. Each spectrum was interpolated to 1 nm intervals by the Matlab function `interp1` with the cubic spline option. Where the original wavelength range was smaller, for example, with no data provided below 400 nm or above 700 nm, the reflectance was extrapolated smoothly. X, Y, Z triplets corresponding to all the reflectance spectra were calculated in both the 1931 and 2015 systems at 1 nm wavelength intervals by the standard triple-product method, i.e., multiplying the reflectance spectrum of the sample by the spectral power distribution of the D65 illuminant and by the tristimulus functions of the CIE 2° Standard Observer, and summing over the visible wavelength range. Spectral observer data was obtained from cvrl.org. Errors were calculated as ΔE^*_{ab} from the 1976 CIELAB formulae based on the 1931 X, Y, Z values. The X, Y, Z values were normalized to $Y_w = 100$, with the D65 white point = [95.04, 100.00, 108.86] for the 1931

system and [94.76, 100.00, 107.54] for the 2015 system. All calculations were performed in Matlab over the wavelength range [380–780] nm, using data vectors with 401 elements.

A regression procedure was employed to find the 3×3 matrix $\mathbf{M}$ that would minimize error in the X, Y, Z domain over the full dataset. With the observation vectors as columns, this can be expressed in right-matrix form as follows. (For left-matrix form with row vectors, the transpose should be used.)

$$\begin{bmatrix} X_{1931} & Y_{1931} & Z_{1931} \end{bmatrix} = \begin{bmatrix} X_{2015} & Y_{2015} & Z_{2015} \end{bmatrix} \mathbf{M} \quad (1)$$

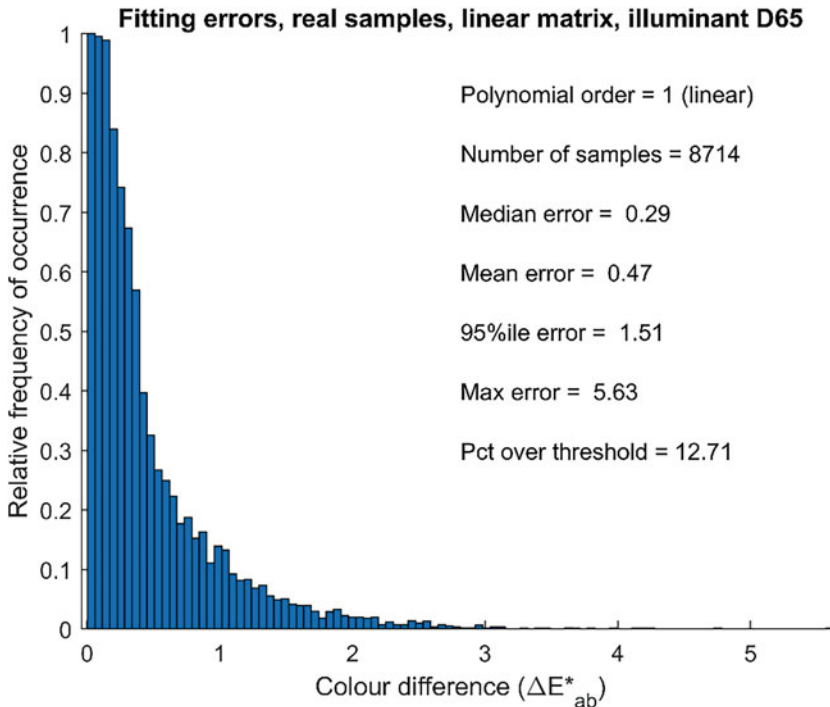
A regression fitting over the full set of 8714 real samples yields the matrix $\mathbf{M}_R$ (with values shown here rounded to four decimal places):

$$\mathbf{M}_R = \begin{bmatrix} 0.9815 & -0.0465 & -0.0590 \\ 0.0099 & 1.0629 & 0.0713 \\ 0.0104 & -0.0181 & 1.0028 \end{bmatrix} \quad (2)$$

The inverse matrix is:

$$\mathbf{M}_R^{-1} = \begin{bmatrix} 1.0178 & 0.0455 & 0.0567 \\ -0.0088 & 0.9393 & -0.0673 \\ -0.0108 & 0.0165 & 0.9954 \end{bmatrix} \quad (3)$$

It is convenient to define the transform in Eq. (1) in the direction from 2015 to 1931, so that the fitting errors can be calculated as color differences in CIE $L^*a^*b^*$, based on the 1931 system. The distribution of color errors is shown in Fig. 2, with mean of $0.47 \Delta E^*_{ab}$ and 12.7% of errors above the visibility threshold, assuming the just-noticeable difference (JND) = 1.0. The performance can be improved slightly by fitting with second- and third-order polynomials.



Transformation of Color Matching Functions, Fig. 2 Distribution of color errors in fitting of real samples, expressed in ΔE^*_{ab}

Fitting with Synthetic Reflectance Spectra

To simulate the widest possible range of reflectance spectra that might be encountered in real environments, a large dataset of 800 K synthetic spectra was generated (see separate article). They were constrained to be continuous, single-valued functions of wavelength at 1 nm intervals over the range 380–780 nm, with a maximum slope (first derivative) not exceeding $\pm 0.024 \text{ nm}^{-1}$ and probability density function of the slope similar to that of real spectra. The gradient values can be related to the cross-sectional area (convex hull) of the color gamut. Whereas the real samples fit within cuboidal bounds of approximately $2 < L^* < 98$, $-61 < a^* < 68$ and $-56 < b^* < 103$, the synthetic spectra fill a larger volume of $0 < L^* < 98$, $-118 < a^* < 115$ and $-105 < b^* < 123$. They are more demanding, therefore, as a test of the robustness of any color transformation.

A regression fitting over the full set of synthetic spectra, again with the D65 illuminant, yields the matrix $\mathbf{M}_S$:

$$\mathbf{M}_S = \begin{bmatrix} 0.9804 & -0.0435 & -0.0739 \\ 0.0111 & 1.0593 & 0.0884 \\ 0.0101 & -0.0171 & 0.9982 \end{bmatrix} \quad (4)$$

The inverse matrix is:

$$\mathbf{M}_S^{-1} = \begin{bmatrix} 1.0187 & 0.0430 & 0.0716 \\ -0.0098 & 0.9423 & -0.0842 \\ -0.0104 & 0.0157 & 0.9997 \end{bmatrix} \quad (5)$$

The error distribution has a longer tail than for the real reflectance spectra, with mean of $0.80 \Delta E^*_{ab}$ and 24.3% of errors above the visibility threshold. The change is greater for the Z_{2015} tristimulus value. Testing this matrix on the real samples gives a mean error of $0.52 \Delta E^*_{ab}$ and 16.1% of errors above the visibility threshold (Fig. 3). Hence, fitting to the wider gamut of synthetic reflectance spectra results in a somewhat poorer performance of the matrix on the real reflectance samples.

Transformation Under Illuminant E

The calculations above were all made for Illuminant D65 (standard daylight). Repeating for Illuminant E (equi-energy) with no normalization gives the following transformation matrices:

$$\mathbf{M}_{RE} = \begin{bmatrix} 0.9287 & -0.0451 & -0.0525 \\ 0.0096 & 1.0064 & 0.0664 \\ 0.0077 & -0.0165 & 0.9360 \end{bmatrix}$$

$$\mathbf{M}_{RE}^{-1} = \begin{bmatrix} 1.0758 & 0.0492 & 0.0569 \\ -0.0096 & 0.9921 & -0.0709 \\ -0.0091 & 0.0170 & 1.0666 \end{bmatrix} \quad (6)$$

For regression on the synthetic spectra:

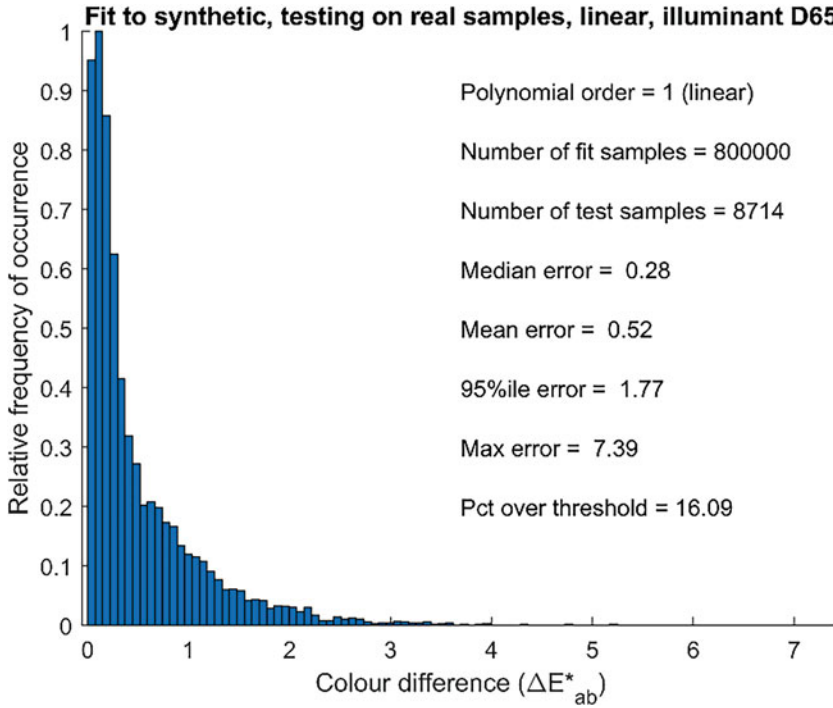
$$\mathbf{M}_{SE} = \begin{bmatrix} 0.9275 & -0.0428 & -0.0649 \\ 0.0110 & 1.0033 & 0.0821 \\ 0.0073 & -0.0154 & 0.9309 \end{bmatrix}$$

$$\mathbf{M}_{SE}^{-1} = \begin{bmatrix} 1.0771 & 0.0470 & 0.0709 \\ -0.0111 & 0.9949 & -0.0885 \\ -0.0086 & 0.0161 & 1.0722 \end{bmatrix} \quad (7)$$

The distribution of color errors for real samples has mean $0.48 \Delta E^*_{ab}$ and 13.5% of errors above the visibility threshold. For synthetic samples, the mean is 0.83 with 25.3% above threshold. Testing the synthetic-fitted matrix with real samples, the mean is 0.54 with 17.1% above threshold.

Conversion of LMS to XYZ

The Stockman-Sharpe 2° cone fundamentals were downloaded from cvrl.org (1 nm linear data). They are defined at 1 nm intervals over the range [390, 780] nm, and each curve is normalized to a maximum value of 1.0 at the peak. In this form, they represent three independent sensors, each of which is univariant with sensitivity dependent on wavelength.



Transformation of Color Matching Functions, Fig. 3 Distribution of color errors in testing of real samples with matrix fitted to synthetic samples

In 2015, the CIE defined tristimulus functions, or color matching functions (CMFs), as a linear combination of the cone fundamentals:

$$\begin{bmatrix} \overline{x}_{2015}(\lambda) & \overline{y}_{2015}(\lambda) & \overline{z}_{2015}(\lambda) \end{bmatrix} = \begin{bmatrix} \overline{l}(\lambda) & \overline{m}(\lambda) & \overline{s}(\lambda) \end{bmatrix} \mathbf{F} \quad (9)$$

where:

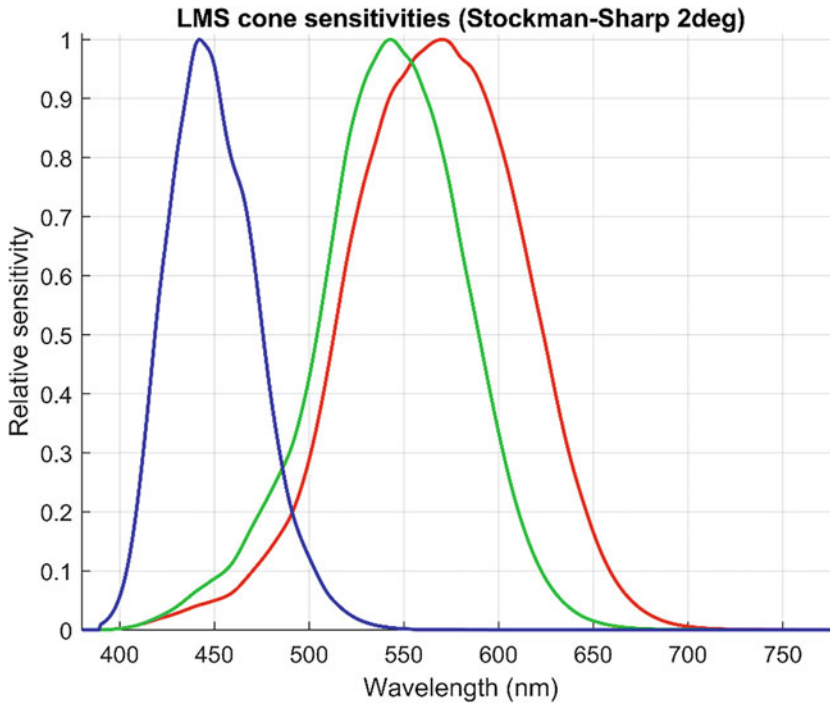
$$\begin{aligned} \overline{x}(\lambda) &= 1.94735469\overline{l}(\lambda) - 1.41445123\overline{m}(\lambda) \\ &\quad + 0.36476327\overline{s}(\lambda) \\ \overline{y}(\lambda) &= 0.68990272\overline{l}(\lambda) + 0.34832189\overline{m}(\lambda) \\ \overline{z}(\lambda) &= 1.93485343\overline{s}(\lambda) \end{aligned} \quad (8)$$

where $\overline{l}(\lambda)$, $\overline{m}(\lambda)$, $\overline{s}(\lambda)$ are the Stockman & Sharpe [5] 2-deg cone fundamentals, adopted by the CIE in 2006. The coefficients were chosen to make the values of the CMFs always positive (i.e., no negative lobes) and to equalize the integrals (i.e., areas under the curves) of all three functions for an equal energy white.

With the observation vectors as columns, Eq. (8) can be expressed in right matrix form as:

$$\mathbf{F} = \begin{bmatrix} 1.94735469 & 0.68990272 & 0 \\ -1.41445123 & 0.34832189 & 0 \\ 0.36476327 & 0 & 1.93485343 \end{bmatrix} \quad (10)$$

Computation was performed as above, calculating cone responses and tristimulus values at 1 nm intervals over the range [380, 780] nm, using the equi-energy white (Illuminant E) and the set of real reflectance spectra. L, M, S values were set to zero in the interval [380, 389] nm. Responses were not normalized. For the 1931 system, the white point is $X_w = Y_w = Z_w = 106.86$, and for the 2015 system $X_w = Y_w = Z_w = 113.04$. For LMS, the integrated responses to equi-energy



Transformation of Color Matching Functions, Fig. 4 Long, medium, and short (LMS) cone fundamentals defined by Stockman and Sharpe

white are not equal: $[L_w, M_w, S_w] = [115.98, 94.82, 58.42]$.

Regression was applied to find the best-fitting matrix between the corresponding sets of triplet values $[L, M, S]$ and $[X, Y, Z]_{2015}$. The transformation matrix resulting from the regression is:

$$\mathbf{F}_{\text{RE}} = \begin{bmatrix} 1.9474 & 0.6899 & 0 \\ -1.4145 & 0.3483 & 0 \\ 0.3648 & 0 & 1.9349 \end{bmatrix}$$

$$\mathbf{F}_{\text{RE}}^{-1} = \begin{bmatrix} 0.2106 & -0.4171 & 0 \\ 0.8551 & 1.1773 & 0 \\ -0.0397 & 0.0786 & 0.5168 \end{bmatrix} \quad (11)$$

$\mathbf{F}_{\text{RE}}$ is identical to $\mathbf{F}$, apart from the number of decimal places, proving that the regression method works correctly. The same result is obtained for regression over the synthetic spectra and with other illumination sources.

We now wish to convert L, M, S to the conventional CIE 1931 tristimulus values X, Y, Z . The relationship can be expressed as:

$$\begin{bmatrix} X_{1931} & Y_{1931} & Z_{1931} \end{bmatrix} = \begin{bmatrix} L & M & S \end{bmatrix} \mathbf{G} \quad (12)$$

Regression fitting over the set of real samples for equi-energy illumination yields the matrix:

$$\mathbf{G}_{\text{RE}} = \begin{bmatrix} 1.8151 & 0.6064 & -0.0565 \\ -1.3102 & 0.4144 & 0.0974 \\ 0.3537 & -0.0483 & 1.7919 \end{bmatrix}$$

$$\mathbf{G}_{\text{RE}}^{-1} = \begin{bmatrix} 0.2663 & -0.3863 & 0.0294 \\ 0.8491 & 1.1664 & -0.0367 \\ -0.0297 & 0.1077 & 0.5513 \end{bmatrix} \quad (13)$$

Regression fitting over the set of synthetic samples for equi-energy illumination yields:

$$\begin{aligned} \mathbf{G}_{SE} &= \begin{bmatrix} 1.8138 & 0.6088 & -0.0697 \\ -1.3081 & 0.4100 & 0.1203 \\ 0.3524 & -0.0455 & 1.7775 \end{bmatrix} \mathbf{G}_{SE}^{-1} \\ &= \begin{bmatrix} 0.2642 & -0.3882 & 0.0366 \\ 0.8519 & 1.1689 & -0.0457 \\ -0.0306 & 0.1069 & 0.5542 \end{bmatrix} \end{aligned} \quad (14)$$

The distributions of color errors for fitting respectively on real and synthetic spectra under Illuminant E have means of 0.48 and 0.83 ΔE^*_{ab} , and error percentages of 13.5% and 25.4% above the visibility threshold. Testing the matrix on the real samples gives a mean error of 0.54 ΔE^*_{ab} and 17.1% of errors above the visibility threshold. These error distributions and their statistics are similar to those arising from the linear matrices fitting the same spectra for the transformation of X, Y, Z tristimulus values from the CIE 1931 to CIE 2015 system, as described in the previous section.

Substituting from Eq. (1) into Eq. (9):

$$\begin{bmatrix} \overline{x_{1931}}(\lambda) & \overline{y_{1931}}(\lambda) & \overline{z_{1931}}(\lambda) \end{bmatrix} = \begin{bmatrix} \overline{l}(\lambda) & \overline{m}(\lambda) & \overline{s}(\lambda) \end{bmatrix} \mathbf{F} \mathbf{M}_R \quad (15)$$

Comparing Eq. (12), it follows that:

$$\mathbf{G} = \mathbf{F} \mathbf{M} \quad (16)$$

To verify this numerically, for real samples under equi-energy illumination we have from Eqs. (6), (11) and (13):

$$\begin{aligned} \mathbf{M}_{RE} &= \begin{bmatrix} 0.9287 & -0.0451 & -0.0525 \\ 0.0096 & 1.0064 & 0.0664 \\ 0.0077 & -0.0165 & 0.9360 \end{bmatrix} \mathbf{F}_{RE} \\ &= \begin{bmatrix} 1.9474 & 0.6899 & 0 \\ -1.4145 & 0.3483 & 0 \\ 0.3648 & 0 & 1.9349 \end{bmatrix} \end{aligned}$$

$$\mathbf{M}_{RE} = \begin{bmatrix} 1.8151 & 0.6064 & -0.0565 \\ -1.3102 & 0.4144 & 0.0974 \\ 0.3537 & -0.0483 & 1.7919 \end{bmatrix}$$

$$\mathbf{M}_{RE} * \mathbf{M}_{RE} = \begin{bmatrix} 1.8151 & 0.6065 & -0.0564 \\ -1.3103 & 0.4143 & 0.0974 \\ 0.3537 & -0.0484 & 1.7919 \end{bmatrix}$$

Transformation Under Illuminant D65

The calculations above for transformation of L, M, S to X, Y, Z were all made for Illuminant E (equi-energy) with no normalization. Repeating for Illuminant D65 (standard daylight) gives different transform matrices. Without normalization, the white responses are:

$$(1931) [X_w, Y_w, Z_w] = [100.43, 105.67, 115.03]$$

$$(2015) [X_w, Y_w, Z_w] = [105.85, 111.71, 120.13]$$

$$(LMS) [L_w, M_w, S_w] = [113.04, 96.81, 62.09]$$

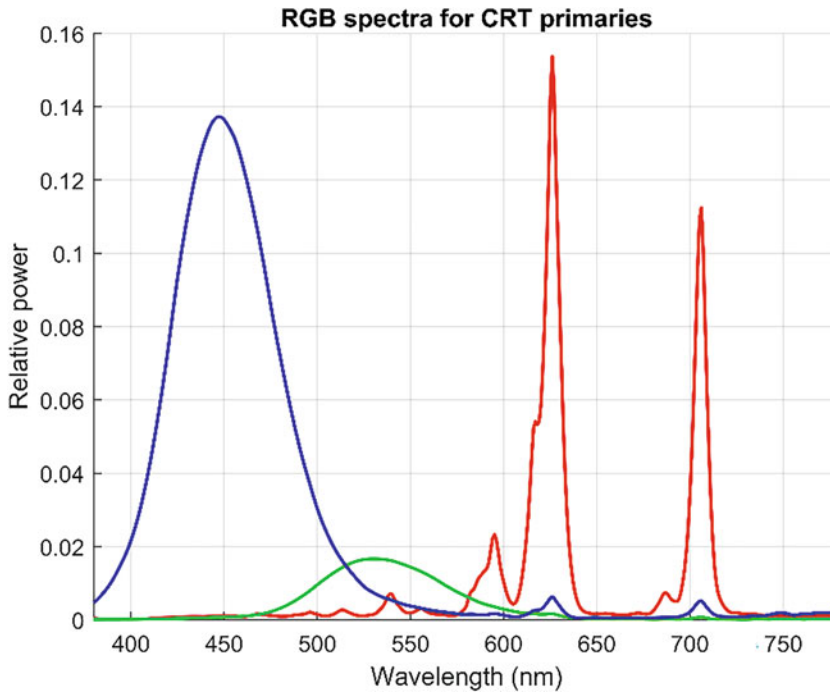
Regression fitting over the set of real samples under D65 yields:

$$\begin{aligned} \mathbf{G}_{R65} &= \begin{bmatrix} 1.8145 & 0.6080 & -0.0623 \\ -1.3100 & 0.4124 & 0.1025 \\ 0.3578 & -0.0491 & 1.8150 \end{bmatrix} \mathbf{G}_{R65}^{-1} \\ &= \begin{bmatrix} 0.2653 & -0.3874 & 0.0310 \\ 0.8499 & 1.1672 & -0.0368 \\ -0.0293 & 0.1080 & 0.5439 \end{bmatrix} \end{aligned} \quad (17)$$

Regression fitting over the set of synthetic samples under D65 yields:

$$\begin{aligned} \mathbf{G}_{S65} &= \begin{bmatrix} 1.8133 & 0.6112 & -0.0784 \\ -1.3082 & 0.4072 & 0.1280 \\ 0.3567 & -0.0463 & 1.8014 \end{bmatrix} \mathbf{G}_{S65}^{-1} \\ &= \begin{bmatrix} 0.2626 & -0.3898 & 0.0391 \\ 0.8531 & 1.1700 & -0.0460 \\ -0.0301 & 0.1072 & 0.5462 \end{bmatrix} \end{aligned} \quad (18)$$

The distributions of color errors for fitting respectively on real and synthetic spectra under D65 have means of 0.47 and 0.80 ΔE^*_{ab} , and error percentages of 12.7% and 24.3% above the visibility threshold, all of which are slightly smaller



Transformation of Color Matching Functions, Fig. 5 Spectral power distribution of the primaries of a CRT display [4]

than for Illuminant E. Testing the matrix on the real samples gives a mean error of $0.52 \Delta E^*_{ab}$ and 16.1% of errors above the visibility threshold.

The computation procedure above could also be applied for other observers, for example, individuals whose color matching functions differ from the CIE Standard Observer, or for a camera or scanner or other instrument with known spectral sensitivity in each channel. They could also be applied for any illumination spectrum, from which it would follow that a best-fitting matrix could be found to transform the tristimulus values from any illuminant A to illuminant B, for a given observer. This would provide the basis for prediction of chromatic adaptation between differing illumination of known spectra.

Transforming Stimuli from Displays

When colors are produced by a display, the scenario is different from above because the

visual stimuli are emitted by three specific primaries, rather than being reflected from a surface illuminated by some source of light. For a CRT display, these are determined by the characteristic R, G, B phosphor spectra, such as those in Fig. 5.

The transformation from linear light R, G, B to tristimulus values X, Y, Z can be represented as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathbf{RGB_XYZ} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (19)$$

where X_r is the contribution to X of the red primary, calculated as $X_r = \sum R(\lambda)X(\lambda)$, etc.

Similarly, the transformation from linear light R, G, B to cone responses L, M, S can be represented as:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} L_r & L_g & L_b \\ M_r & M_g & M_b \\ S_r & S_g & S_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathbf{RGB_LMS} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (20)$$

from which it follows that:

$$\begin{aligned} \mathbf{RGB_XYZ}^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= \begin{bmatrix} R \\ G \\ B \end{bmatrix} \\ &= \mathbf{RGB_LMS}^{-1} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \end{aligned} \quad (21)$$

Hence:

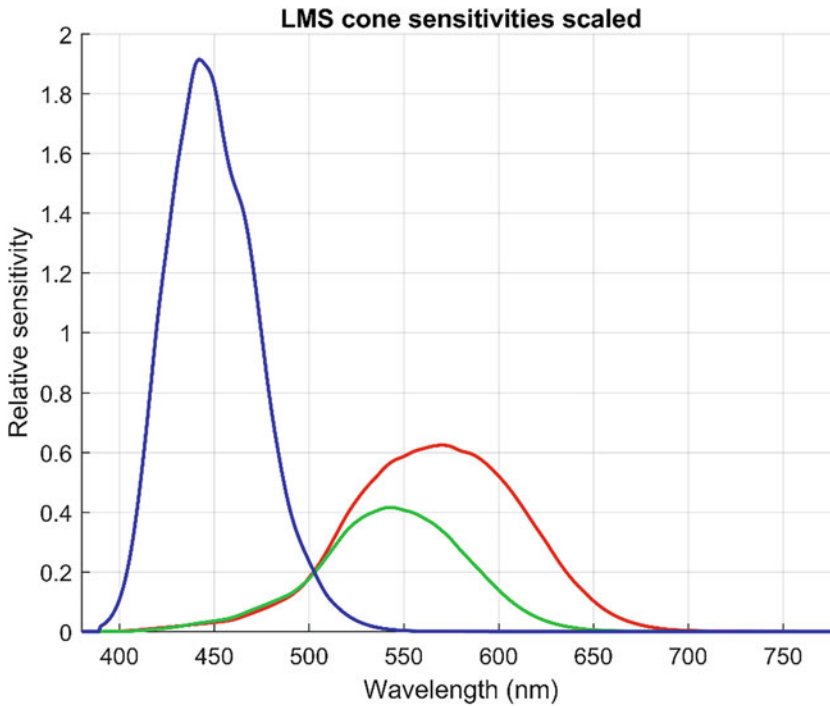
$$\begin{aligned} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= \mathbf{RGB_XYZ} \cdot \mathbf{RGB_LMS}^{-1} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \\ &= \mathbf{LMS_XYZ} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \end{aligned} \quad (22)$$

Using the same scaling of X, Y, Z and L, M, S as in the previous sections, the matrices are computed as:

$$\begin{aligned} \mathbf{RGB_XYZ} &= \begin{bmatrix} 1.8113 & 0.4821 & 2.1926 \\ 1 & 1 & 1 \\ 0.1098 & 0.1923 & 11.1304 \end{bmatrix}, \\ \mathbf{RGB_LMS} &= \begin{bmatrix} 0.8373 & 0.6075 & 0.6604 \\ 0.2013 & 0.4169 & 0.6141 \\ 0.1090 & 0.1693 & 11.5769 \end{bmatrix} \end{aligned} \quad (23)$$

and so:

$$\begin{aligned} \mathbf{XYZ_LMS} &= \begin{bmatrix} 0.2745 & 0.8466 & -0.0352 \\ -0.3830 & 1.1655 & 0.1032 \\ 0.0100 & -0.0208 & 0.5429 \end{bmatrix}, \\ \mathbf{LMS_XYZ} &= \begin{bmatrix} 1.8072 & -1.3061 & 0.3653 \\ 0.5948 & 0.4253 & -0.0423 \\ -0.0105 & 0.0404 & 1.8335 \end{bmatrix} \end{aligned} \quad (24)$$



Transformation of Color Matching Functions, Fig. 6 LMS cone fundamentals expressed in units of luminance

Note that the original cone spectral sensitivities were scaled to unity (Fig. 4). To express them in luminance units, the scaling factors given by Stockman and Sharpe need to be applied. The cone response functions are multiplied by the respective scaling factors:

$$k_L = 0.624899, k_M = 0.416602, k_S = 1.915443 \quad (25)$$

The scaled cone response functions then appear as in Fig. 6.

The transform and its inverse in Eq. (24) change to the following:

$$\begin{aligned} \mathbf{XYZ}_{\text{LMS}} &= \begin{bmatrix} 0.1716 & 0.5290 & -0.0220 \\ -0.1595 & 0.4855 & 0.0430 \\ 0.0191 & -0.0399 & 1.0399 \end{bmatrix}, \\ \mathbf{LMS_XYZ} &= \begin{bmatrix} 2.8919 & -3.1352 & 0.1907 \\ 0.9518 & 1.0208 & -0.0221 \\ -0.0167 & 0.0969 & 0.9572 \end{bmatrix} \end{aligned} \quad (26)$$

These equations are different from Eqs. (13) and (14), because the viewing environment involves color stimuli generated by a CRT display. All stimuli in this case are weighted combinations of the spectra of the three primaries, modulated by the signal level. It is assumed that the device is linear and additive, with no cross-talk between channels and with negligible ambient illumination, thus making possible the simple formulation in Eqs. (19) and (20).

One could imagine a different scenario in which the display white was used as the light source to illuminate a series of reflective samples with known reflectance spectra, but in this case the light reaching the eye for each sample would consist of the display white spectrum multiplied by the reflectance spectrum of the surface. This would lead through regression to a different transformation matrix. A display with white balance matched to D65 is metameric to a daylight spectrum correlated to 6500 K.

The method given above in Eqs. (19) to (26) could be applied to any display technology, such as LCD or LED, provided that the device is additive and free from cross-talk. All that is required is to measure the spectra of the three primaries.

Cross-References

- CIE Color-Matching Functions
- CIE Cone Fundamentals
- CIE Standard Illuminant D65
- CIE Tristimulus Values

References

1. CIE: Publication 15.2 Colorimetry, 2nd edn, (1986)
2. CIE: Publication 015:2018 Colorimetry, 4th edn, (2018)
3. CIE: Fundamental Chromaticity Diagram with Physiological Axes – Part 1, Technical Report 170–1. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
4. Golz, J., MacLeod, D.I.: Colorimetry for CRT displays. *J. Opt. Soc. Am. A*. **20**(5), 769–781 (2003) The spectral data of the display primaries is at: <https://www.uni-kiel.de/psychologie/golz/publications/2003a/AveragePhosphors.dat>
5. Stockman, Sharpe: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Res.* **40**, 1711–1737 (2000)

Transient Imaging

Adrian Jarabo

Graphics and Imaging Lab, Universidad de Zaragoza, I3A, Zaragoza, Spain

Synonyms

Femto-photography; Light-in-flight imaging; Ultrafast Imaging

Definition

Transient imaging refers to the set of techniques within the field of computational imaging that focus on capturing and analyzing the propagation of ultrashort light pulses and its interaction with matter, at frame rates comparable to the speed of light. It allows to capture the optical impulse response of a scene as a function of time. This has multiple applications in scene understanding and computer vision, including range imaging, non-line-of-sight reconstruction, vision through turbid media, BRDF estimation, or light-in-flight imaging.

Transient State Versus Steady State

One of the most general assumptions in computer vision and scene understanding is to consider the speed of light to be infinite, therefore assuming light transport in steady-state. This is a reasonable assumption since most of the existing imaging hardware is very slow compared to the speed of light. Even in the case of high-speed cameras, capable to capture up to a few thousands of frames per second, light travels tens of kilometers in each frame (Fig. 1, left). With the emergence of ultrafast imaging devices, effective exposure times of nano- or picoseconds have been made possible. At such exposure times, light travels a few centimeters per frame, effectively breaking the assumption of steady state of light, requiring to account for the time-resolved nature of light transport. Light $L(x, t)$ incoming at the camera's pixel x is a function defined on the temporal domain

t (Fig. 1). This time-resolved radiance relates with its steady-state counterpart $L_s(x)$ as:

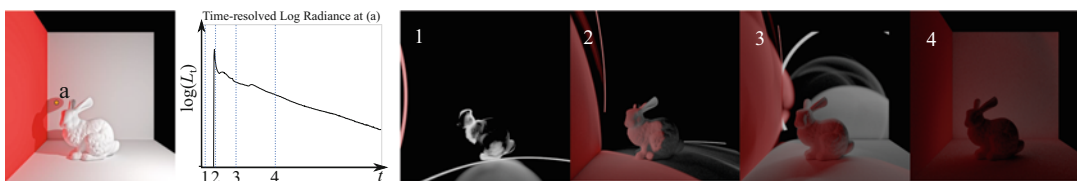
$$L_s(x) = \int_{-\infty}^{\infty} L_t(x, t) dt. \quad (1)$$

Breaking the steady-state assumption has led to many promising applications in computer vision and scene understanding, including robust range imaging, non-line-of-sight reconstruction, vision through turbid media, or reflectance recovery.

Historical Development

Ultrafast imaging of impulse illumination in macroscopic scenes was first proposed in the 1970s by Abramson [1]; he proposed to use holography, achieving femtosecond temporal resolution. However, Abramson's technique was based on light coherence. Therefore, it focused in direct illumination, resulting in the loss of indirect illumination. Ultrafast imaging of direct illumination has been an active field on its own for decades, given its importance for range imaging for, e.g., robotics, remote sensing, or 3D scanning. In this context, two main technologies stand out: LiDAR (*light detection and ranging*), which measures distance by taking into account the light time of flight from an ultrashort laser pulse to the camera; and *time-of-flight* (ToF) cameras that use coded illumination and sensor for sensing depth. These two technologies focused on depth imaging, rather than capturing light in flight; more details on range sensors can be found in [5].

The first system capable of imaging light in motion at large spatial and temporal resolution



Transient Imaging, Fig. 1 Synthetic example of time-resolved light transport. In steady state (left), all light has finished propagating along the scene. In transient state, however, the radiance incoming at pixel is a function of

time, and light takes several nanoseconds to finishing propagating along the scene. The plot shows the temporal profile of radiance in pixel (a), while the individual frames are shown as blue dashed lines in the plot

was Velten et al.'s *femto-photography* [14]. They proposed to combine an ultrafast sensor with femtosecond laser, to capture videos at picosecond resolution of light in motion, including indirect illumination or multiple scattering. Inspired by this system, many other works have been proposed, either developing cheaper, more robust capture systems [3, 6], reducing the amount of measurements by single-shot transient imaging using compressed sensing [2], or significantly increasing the temporal resolution by using optical time-stretching via dispersive optics [11].

In the following, the different approaches for capturing transient light transport are described, as well as some of the most promising applications of this technology. An in-depth discussion can be found in recent surveys on the topic [8].

Capturing Light in Flight

Imaging light in flight is equivalent to capture the impulse response of a scene along the temporal domain. The impulse response of a scene is a linear operator that relates the scene illumination with the captured image, via the light transport equation:

$$\mathbf{i}(t) = \int_{-\infty}^{\infty} \mathbf{T}(\tau) \mathbf{p}(t - \tau) d\tau = (\mathbf{T} * \mathbf{p})(t), \quad (2)$$

where $\mathbf{i}(t)$ is the spatio-temporal image, $\mathbf{p}(t)$ is the time-resolved illumination function, and $\mathbf{T}(t)$ is the light transport matrix describing the light transport with a time-of-flight of exactly t . The second equality represents the convolution in the temporal domain between $\mathbf{T}$ and $\mathbf{p}$. The ideal case for obtaining $\mathbf{T}(t)$ in a scene is to use a delta illumination in the temporal domain. However, this is in general expensive, and different cheaper approaches have been proposed for capturing $\mathbf{T}(t)$. These systems can be categorized in three main categories:

1. Straight Temporal Recording

This is the conceptually simplest approach and consists on using pulsed illumination and ultrafast imaging hardware to capture Eq. (2).

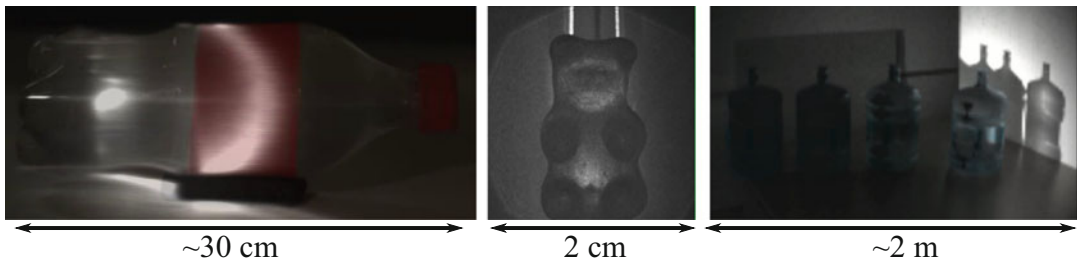
Several different optical systems have been proposed, including streak-cameras [14], single photon avalanche diodes (SPAD) [3], or dispersive optics combined with white-light lasers [11]. These systems range in terms of cost, temporal resolution, operating time, or spatial resolution: For example, while SPADs are cheap, they have limited spatio-temporal resolution; streak-cameras, on the other hand, are expensive, difficult to operate, and need long capture times but have large spatial resolution, and effective exposure times in the order of 2 picoseconds.

2. Interferometry

Interferometry was the first technology for imaging light in motion [1], although its limited range of applicability has made its scope relatively narrow. The key idea is to leverage electromagnetic interference of coherent light to disambiguate between slightly different path lengths. This results into a very high temporal resolution (at the femtosecond scale), at the price of working in small scenes and losing incoherent light such as indirect illumination.

3. Phase-Based Time-of-Flight Cameras

The last type of systems used for transient imaging base on phase-based time-of-flight (P-ToF) cameras is called correlation-based time-of-flight (C-ToF) or simply ToF cameras [5]. These are the cheapest systems of the three and were designed for low-cost time-of-flight-based range imaging with high spatial resolution. P-ToF cameras cross-correlate emitted time-modulated light and the sensor response with (generally) a periodic continuous wave. This correlation encodes the impulse response of a pixel modulated at the sensor. By using a small set of measurements with different phase-shifts, it allows to disambiguate light time-of-flight. Beyond range imaging, Heide and colleagues [6] demonstrated that by taking several measurements with different modulation frequencies and phase shifts, it was possible to reconstruct transient images using optimization, leading to low-budget transient imaging using easy-to-operate hardware. The main drawback of these systems is the



Transient Imaging, Fig. 2 Example frames of light-in-flight visualizations captured by using (from left to right) straight temporal recording (femtophotography [14]), interferometry [4], and phase-based ToF cameras [6]. Each device has its different capabilities and range of

applicability, being able to capturing scenes with different scales, and imposing a temporal resolution of up to around 1–2 ps (straight recording), 30 fs (interferometry), and 1 ns (ToF cameras). Full visualizations of the captured data can be found in the respective sources

significantly lower temporal resolution, compared with the two alternative approaches.

Figure 2 shows example frames of light-in-flight visualizations of captured transient data using each of these three approaches.

Applications

Time-resolved data has opened up a wide set of applications. Long-standing problems in computer vision like depth recovery or vision through turbid media, to name a few, have received renewed attention in light of this new data. Novel applications on scene understanding have additionally emerged thanks to the additional information encoded in time-of-flight data, including robust range imaging in the presence of interreflections and scattering media, material estimation, global illumination separation, light-in-flight imaging, or non-line-of-sight reconstruction of hidden scenes. The most relevant applications are described in the following:

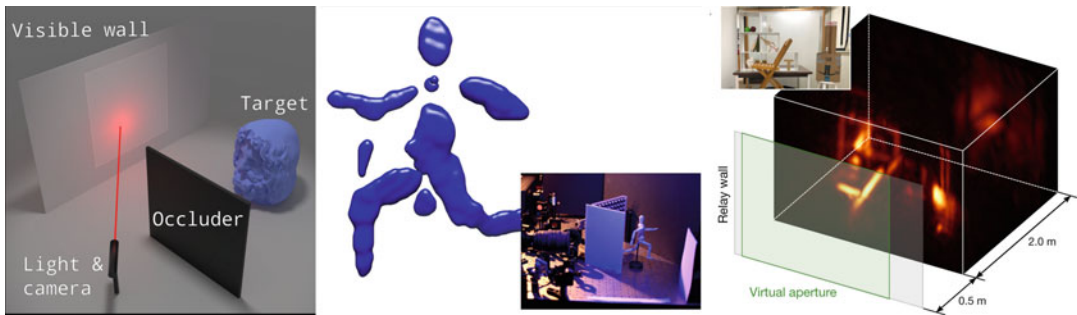
Range Imaging

Range imaging aims to recover the depth information from a scene and is a key problem in computer vision. While several approaches have been proposed to tackle the problem (including stereo and multiview reconstruction or structured illumination), transient-based range imaging is one of the most successful approaches. Intuitively, the idea is to measure the time (and consequently

the distance traveled) that takes a light pulse to travel from a calibrated emitter source to the sensor, so that the first incoming response of the scene encodes the depth at a pixel. This is the main application for transient image for decades, in particular for LIDAR or time-of-flight sensors [5]. However, these techniques are sensitive to errors produced by **multipath interference** (MPI), where light from multiple light paths affects the range measurement of direct illumination, resulting in errors. This type of errors might be problematic in the case of indirect illumination from nearby geometry but is completely pathological for imaging through turbid media, where MPI dominates. Capturing the full temporal response has allowed to disambiguate direct and indirect illumination, as well as the contribution of media. This has resulted into significantly more robust depth estimation techniques (e.g., [15]), even in the presence of scattering media using cheap ToF sensors [7].

Material Estimation and Classification

Time-resolved data gives valuable information on the optical properties of surfaces and media, given the particular time profiles of light-matter interactions. For example, fluorescence lifetime imaging allows to measure the reemission response of fluorescent materials even through turbid media and uses this response to detect and classify such materials. The temporal response has been also used to capture the optical properties of translucent objects and use them for classification, based on the reflectance decay in the temporal domain



Transient Imaging, Fig. 3 Left: Canonical setup for NLOS imaging, where a hidden object is recovered from indirect reflections in a visible relay wall. Middle: While this technology has enormous potential in several applications, most methods have managed to capture only small, isolated objects in carefully controlled laboratory

conditions [13]. Right: The applicability of NLOS imaging has significantly increased recently, with the development of *phasor fields* [9], which allow reconstructing large scenes with arbitrary geometrical complexity and materials. The insets show the scenes to be reconstructed

[15]. Finally, the ability of demultiplex light paths along its path length has been used for creating virtual gonireflectometers for capturing materials reflectance.

Non-line-of-Sight Imaging

Non-line-of-sight (NLOS) imaging aims to recover objects that are not directly in the line of sight of the observer [10]. It is generally known as “looking around corners,” although its range of applications goes beyond these scenarios. The key idea is to use the indirect reflections on a secondary visible wall to reconstruct the hidden scene (see Fig. 3, left). It was first demonstrated by the seminal work by Velten and colleagues [13], who used a setup similar to *femto-photography* [14] for reconstructing hidden simple tabletop scenes. The problem has gained significant attention over the last few years, leading to significant advances allowing to use cheaper sensors and demonstrated real-time performance [12]. Most works in this field have assumed very simplistic scenes, with simple isolated objects with small contribution from high-order scattering, and simplistic materials (Fig. 3, middle). Most recently, NLOS imaging was proved to work on more challenging conditions, including scenarios with significant complexity and range, arbitrary materials, and significant MPI, by posing the inverse problem as a virtual forward light transport model [9] (Fig. 3, right). This has allowed to use knowledge

from traditional Fourier optics into the NLOS domain, which might lead to many novel applications in NLOS imaging in general.

Cross-References

- [Global Illumination](#)
- [Photodetector](#)

References

1. Abramson, N.: Light-in-flight recording by holography. *Opt. Lett.* **3**(4), 121–123 (1978)
2. Gao, L., Liang, J., Li, C., Wang, L.V.: Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*. **516**(7529), 74 (2014)
3. Gariepy, G., Krstajić, N., Henderson, R., Li, C., Thomson, R.R., Buller, G.S., Heshmat, B., Raskar, R., Leach, J., Faccio, D.: Single-photon sensitive light-in-flight imaging. *Nat. Commun.* **6**, 6021 (2015)
4. Gkioulekas, I., Levin, A., Durand, F., Zickler, T.: Micron-scale light transport decomposition using interferometry. *ACM Trans. Graph.* **34**(4), 1–14 (2015)
5. Hansard, M., Lee, S., Choi, O., Horaud, R.P.: *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer Science & Business Media, Berlin (2012)
6. Heide, F., Hullin, M.B., Gregson, J., Heidrich, W.: Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph.* **32**(4), 45 (2013)
7. Heide, F., Xiao, L., Kolb, A., Hullin, M.B., Heidrich, W.: Imaging in scattering media using correlation

- image sensors and sparse convolutional coding. *Opt. Express*. **22**(21), 26338–26350 (2014)
8. Jarabo, A., Masia, B., Marco, J., Gutierrez, D.: Recent advances in transient imaging: a computer graphics and vision perspective. *Visual Inform.* **1**(1), 65–79 (2017)
 9. Liu, X., Guillén, I., La Manna, M., Nam, J.H., Reza, S.A., Le, T.H., Jarabo, A., Gutierrez, D., Velten, A.: Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*. **572**(7771), 620–623 (2019)
 10. Maeda, T., Satat, G., Swedish, T., Sinha, L., Raskar, R.: Recent advances in imaging around corners. *arXiv preprint arXiv:1910.05613* (2019)
 11. Nakagawa, K., Iwasaki, A., Oishi, Y., Horisaki, R., Tsukamoto, A., Nakamura, A., Hirose, K., Liao, H., Ushida, T., Goda, K., et al.: Sequentially timed all-optical mapping photography (STAMP). *Nat. Photonics*. **8**(9), 695 (2014)
 12. O'Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*. **555**(7696), 338 (2018)
 13. Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M.G., Raskar, R.: Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 745 (2012)
 14. Velten, A., Wu, D., Jarabo, A., Masia, B., Barsi, C., Joshi, C., Lawson, E., Bawendi, M., Gutierrez, D., Raskar, R.: Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* **32**(4), 1–8 (2013)
 15. Wu, D., Velten, A., O'Toole, M., Masia, B., Agrawal, A., Dai, Q., Raskar, R.: Decomposing global light transport using time of flight imaging. *Int. J. Comput. Vis.* **107**(2), 123–138 (2014)

Transition Metal Ion Colors

Richard J. D. Tilley
Cardiff University, Cardiff, UK

Synonyms

d-Block elements

Definition

Transition metals are d-block elements with partially filled 3d, 4d, and 5d orbitals.

Color Production

The transition metals are (somewhat imprecisely) described as being colored because their compounds frequently take on a characteristic hue (Table 1). For example, Ni^{2+} cations tend to exhibit green coloration, while Cu^{2+} tends to show green-blue tones. Cobalt, as Co^{2+} , has been used since medieval times to color glass deep blue, and small amounts of impurity Cr^{3+} in colorless Al_2O_3 produce the color in ruby gemstones (Fig. 1). All these colors arise from electronic transitions between the ionic ground state and energy levels lying between 1.77 and 3.10 eV above it, giving absorption maxima in the visible wavelength range (400–700 nm). These low-lying energy levels arise from interactions of the d orbitals on the cation with neighboring atoms in a material and are a function of the symmetry of the surroundings [1–3]. The energy levels that occur with the 4d and 5d transition metals are higher in energy than those of the 3d series and are of lesser importance in color terms. In this entry, only the 3d transition metal ions are covered and, in addition, only single dopants are considered. (When two or more transition metal ion impurities are present, an alternative mechanism, charge transfer, can also lead to coloration.)

Electronic Energy Levels

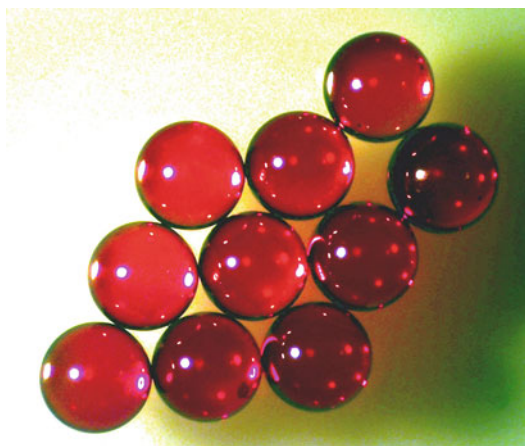
Energy Levels and Terms

The energy levels of a free transition metal cation are dominated by electron–electron repulsion and are described by *term symbols*. A *term* is a set of states which are very similar in energy. Transitions between terms, or, more precisely, the energy levels specified by the term, give rise to the observed line spectrum of the free ion. The terms of an ion can be derived by Russell–Saunders (*LS*) coupling [2, 4]. In this designation, each term is written as ^{2S+1}L where L is a many-electron quantum number describing the total orbital angular momentum of the electrons surrounding the atomic nucleus and S is a many-electron quantum number representing the total electron spin. The superscript $(2S + 1)$ is called the *multiplicity* of the term. The total angular

Transition Metal Ion Colors, Table 1 Typical 3d transition metal ion colors

Ion	Color	Symmetry	Host matrix
Ti ³⁺ d ¹	Purple	Tetrahedral	Silicate glass
V ⁴⁺ d ¹	Red	Tetrahedral	Silicate glass
V ³⁺ d ²	Green Blue	Tetrahedral Octahedral	Silicate glass Al ₂ O ₃
Cr ³⁺ d ³	Green Green Red Violet	Tetrahedral Octahedral Octahedral Octahedral	Silicate glass Be ₃ Al ₂ Si ₆ O ₁₈ (emerald) Al ₂ O ₃ (ruby), TiO ₂ , MgAl ₂ O ₄ KCr(SO ₄) ₃ ·12H ₂ O (chrome alum)
Mn ³⁺ d ⁴	Purple	Tetrahedral	Silicate glass
Mn ²⁺ d ⁵	Yellow Red Green Pink	Tetrahedral Octahedral Octahedral Octahedral	Silicate glass MnCO ₃ MnO MnSiO ₃
Fe ³⁺ d ⁵	Green Yellow	Tetrahedral Octahedral	Silicate glass Al ₂ SiO ₄ (OH) (topaz)
Fe ²⁺ d ⁶	Green Red	Octahedral Cubic	Fe(H ₂ O) ₆ ²⁺ in solution and hydrates Ca ₃ Al ₂ Si ₃ O ₁₂ (garnet)
Co ²⁺ d ⁷	Blue Pink	Tetrahedral Octahedral	CoAl ₂ O ₄ , silicate glass Co(H ₂ O) ₆ ²⁺ in solution and hydrates
Ni ²⁺ d ⁸	Green Yellow	Octahedral Octahedral	Ni(H ₂ O) ₆ ²⁺ in solution and hydrates Al ₂ O ₃ , NiCl ₂
Cu ²⁺ d ⁹	Green Blue	Octahedral ^a Octahedral ^a	Cu ₂ (OH) ₂ (CO ₃) (malachite) Cu(H ₂ O) ₆ ²⁺ in solution and hydrates

^aNote that the octahedral coordination in Cu²⁺ compounds is often distorted with (usually) two long axial bonds and four shorter equatorial bonds, to the extent that the coordination polyhedron is often described as square planar



Transition Metal Ion Colors, Fig. 1 Synthetic ruby spheres consisting of colorless Al₂O₃ doped with trace amounts of Cr³⁺

momentum quantum number L is given a *letter symbol*, S ($L = 0$), P ($L = 1$), D ($L = 2$), and F ($L = 3$), and thereafter alphabetically, omitting J. The energies of the terms must be determined

by quantum mechanical calculations, except for that of the ground state, which is given by Hund's second rule: the ground state is the term with the highest multiplicity and, if more than one term of the same multiplicity is present, by that with the highest L value. The low-energy terms of free 3d transition metal cations are given in Table 2.

Crystal Field Splitting

When a cation is placed into a crystal, the d electrons interact with the surrounding anions and give rise to new energy states. The simplest conceptual model that describes the formation of the color-producing energy levels is called *crystal field theory* [5, 6]. In this approach, the electrostatic effect of neighboring anions upon the energies of the cationic d orbitals is considered in terms of surrounding point charges. More comprehensive treatments, which include the effects of covalent bonding, fall into the domain of *ligand field theory* [7, 8], although the two expressions are often used interchangeably.

The d orbitals on a 3d transition metal cation can be described as pointing along or between a set of x-, y-, and z-axes (Fig. 2). The orbitals directed between the axes are the d_{xy} , d_{yz} , and d_{xz} set, and

Transition Metal Ion Colors, Table 2 Low-energy terms of free 3d transition metal cations

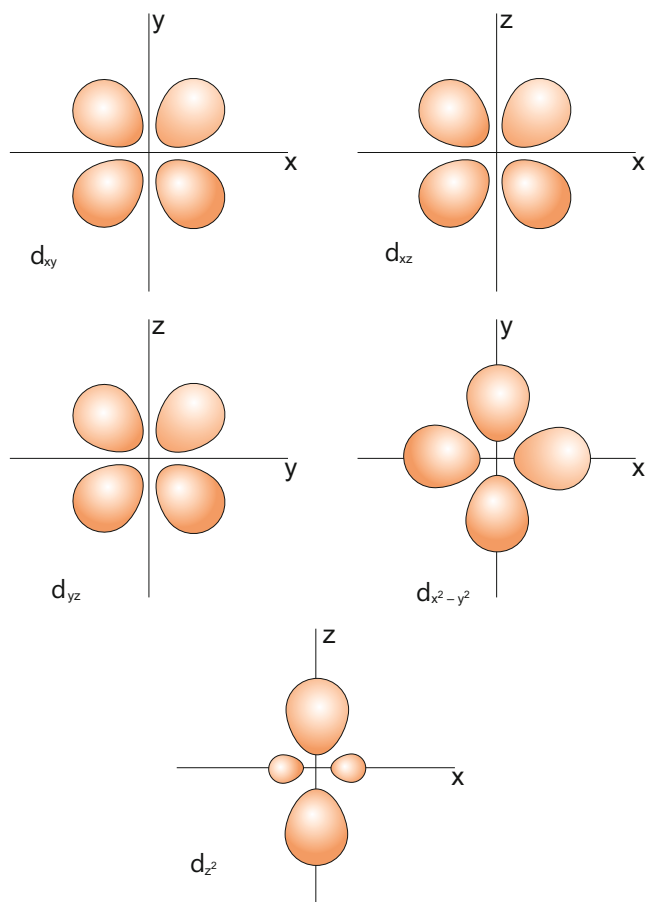
Ion	Free ion terms ^a
Ti ³⁺ V ⁴⁺ d ¹	²D
V ³⁺ d ²	³F ³ P ¹ D ¹ G
Cr ³⁺ d ³	⁴F ⁴ P ² G ² P
Mn ³⁺ Cr ²⁺ d ⁴	⁵D
Mn ²⁺ Fe ³⁺ d ⁵	⁶S ⁴ G ⁴ P ⁴ D ⁴ F
Fe ²⁺ d ⁶	⁵D
Co ²⁺ d ⁷	⁴F ⁴ P ² G ² P
Ni ²⁺ d ⁸	³F ³ P ¹ D ¹ G
Cu ²⁺ d ⁹	²D

^aGround state term **bold**

those pointing along the axes are the $d_{x^2-y^2}$ and d_{z^2} pair. If electron–electron interactions are ignored, all of these orbitals are degenerated (have the same energy). However, this is not true when the ion is placed into a crystal because of the interaction (most easily imagined as repulsion) between the electrons on the surrounding ions and the electrons occupying d orbitals. If the surrounding charges are distributed evenly over the surface of a sphere, the five d orbitals are still energetically degenerated, although at a higher energy by an amount E_0 . In a crystal the charges are not smeared out but arranged in a pattern. This removes the degeneracy and gives rise to a new set of energy levels (Fig. 3). The pattern of this *crystal field splitting* depends upon the symmetry of the surrounding anions.

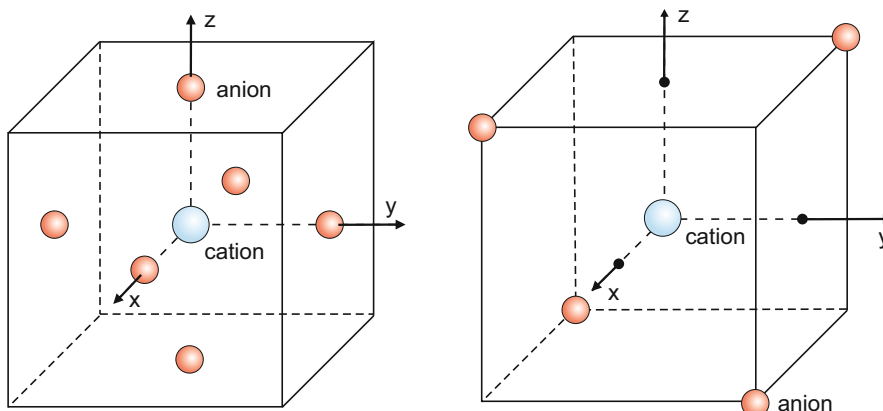
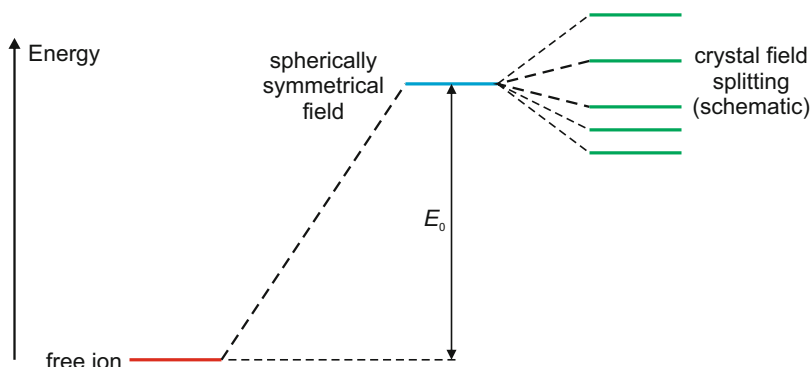
The two most important geometries to consider, especially for oxide pigments and ceramics, are octahedral and tetrahedral coordination

Transition Metal Ion Colors, Fig. 2 Electron density lobes of the d orbitals



Transition Metal Ion

Colors, Fig. 3 Change of d-orbital energies when a free ion is introduced into a crystal (schematic)



Transition Metal Ion Colors, Fig. 4 (a) Octahedral and (b) tetrahedral coordination geometry. A cation is at the cube center. In (a) anions are at the center of each face. In (b) four anions occupy half the cube corners

(Fig. 4a, b). When a cation is surrounded by six anions arranged at the vertices of a regular octahedron, the $d_{x^2-y^2}$ and d_{z^2} orbitals point directly toward the surrounding anions and are raised in energy more than the d_{xy} , d_{xz} , and d_{yz} trio, which point between the anions. This splits the originally degenerated energy levels into two groups. The higher energy state, labeled e_g , consists of two equal energy levels, derived from the $d_{x^2-y^2}$ and d_{z^2} orbitals. The lower energy state, labeled t_{2g} , consists of three equal energy levels, derived from the d_{xy} , d_{xz} , and d_{yz} orbitals (Fig. 5). The energy gap between the t_{2g} orbitals and the e_g orbitals is written Δ or $10Dq$, with the t_{2g} set at $-4Dq$ and the e_g set at $+6Dq$ with respect to the spherically symmetrical situation. When Δ is small, the crystal field is said to be *weak*, and when large it is referred as *strong*. Strong crystal fields arise from ions with multiple charges and relatively short

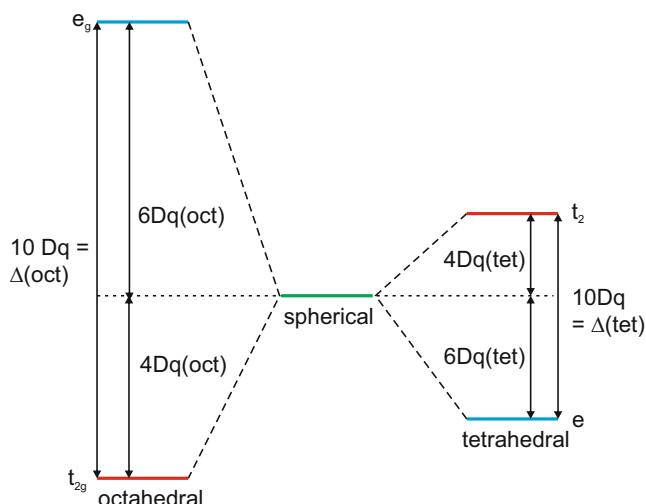
cation–anion spacing, together with contributions due to covalent bonding.

When a cation is surrounded by a tetrahedron of anions, the crystal field splitting found for octahedral coordination is reversed. In this case, the higher energy group, now labeled t_2 , is derived from the d_{xy} , d_{xz} , and d_{yz} orbitals, and the lower energy group, now labeled e , is derived from the $d_{x^2-y^2}$ and d_{z^2} orbitals. The t_2 set is at $+4Dq$ and the e set is at $-6Dq$ with respect to the spherically symmetrical situation (Fig. 5). The magnitude of Δ for tetrahedrally coordinated cations is $4/9$ of that for the same cation when octahedrally coordinated.

Term Splitting

The electronic energy levels of a free transition metal ion arise from electron–electron interactions, and the simple model just outlined must be

Transition Metal Ion Colors, Fig. 5 Splitting of d-orbital energy levels in a crystal field: octahedral and tetrahedral symmetries



Transition Metal Ion Colors, Table 3 Splitting of terms in fields of cubic symmetry

Free ion term	Terms in tetrahedral crystal field	Terms in octahedral crystal field
S	A_1	A_{1g}
P	T_1	T_{1g}
D	E, T_2	E_g, T_{2g}
F	A_2, T_1, T_2	A_{2g}, T_{1g}, T_{2g}
G	A_1, E, T_1, T_2	$A_{1g}, E_g, T_{1g}, T_{2g}$

adapted so as to apply to the free ion terms. It is found that a term may split into several energy states when the cation is subjected to the crystal field. The number of states that arise is a function of the value of L and the symmetry of the surrounding neighboring anions (Table 3). The magnitude of the splitting is a function of the anion–cation separation and the charges on the ions involved. The multiplicity of the term is carried over onto the new states. The energy states in the crystal are given labels that describe the degeneracy of the orbitals: A is a singly degenerate state, E represents a doubly degenerate state, and T represents a triply degenerate state. Different configurations of sets of orbitals with the same degeneracy are labeled with subscripts 1, 2, and so on. Thus a P term gives rise to a T_1 term, while a D term gives rise to a T_2 term. The subscript “g” means a center of orbital symmetry exists. The energy of the new states needs to be calculated

using quantum mechanical methods or determined experimentally from spectra.

Selection Rules

Electron transitions are governed by selection rules that give the *probability* that a transition will occur. Transitions between d orbitals are forbidden by the Laporte selection rule. However, this rule may break down for ions in compounds. The main reason for this is a degree of mixing between s, p, and d orbitals can occur when an ion is not located at a center of symmetry. As s or p to d transitions are allowed, transitions giving rise to color are also allowed, to a degree corresponding to the amount of orbital mixing achieved. Thus ions situated in tetrahedral coordination, which are not at a center of symmetry, show quite strong colors. Ions at the center of a perfect octahedron are at a center of symmetry, and color-producing transitions are forbidden, but in most solids and liquids, thermal agitation of the surroundings and crystal distortions remove the precise symmetry, making such transitions weakly allowed. As a consequence, they are often less intense than those from similar ions in tetrahedral sites.

In addition, transitions are only allowed between states of the same multiplicity, called *spin-allowed transitions*. Transitions between states of differing spin can be weakly allowed in some circumstances, but in general these do not give rise to strong colors.

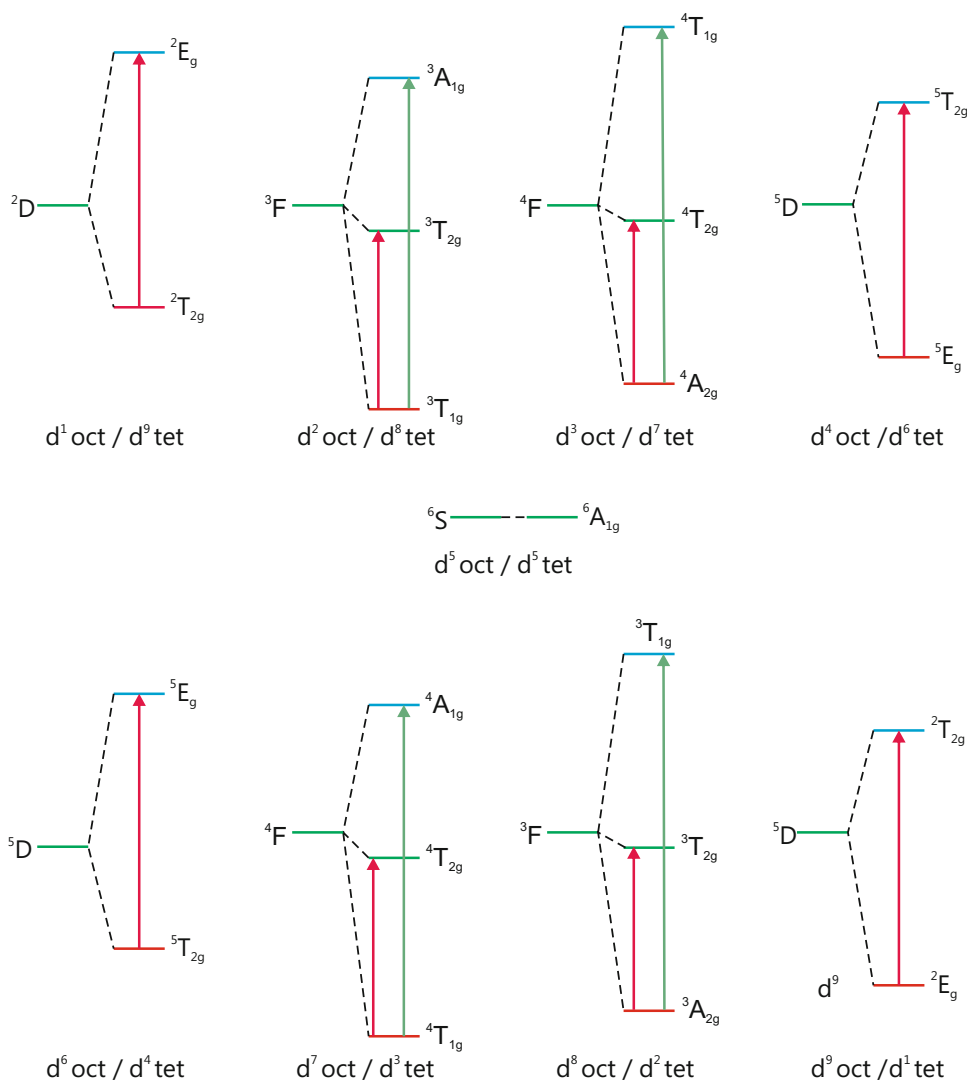
The Color of 3d Transition Metal Ions

Color and Term Splitting

Using Tables 2 and 3, a qualitative energy level diagram can be readily constructed for the splitting of the ground state terms (Fig. 6). It is seen that the arrangement of the energy levels is symmetrical about the d^5 (6S) term. The splitting for a d^n ion in an octahedral field is the same as that for a d^{10-n} ion in a tetrahedral field. Recall, though, that the magnitude of the splitting for cations in a

tetrahedral field is $4/9$ of that in an octahedral field. The colors exhibited by many transition metal ions can now be understood.

Figure 6 indicates that octahedrally coordinated d^5 ions (Fe^{3+} , Mn^{2+}) exhibit no ground state crystal field splitting and are not expected to show any colors. This is so, and compounds of these ions are at best weakly colored due to other factors. The free ion ground state of octahedrally coordinated d^1 ions (Ti^{3+} , V^{4+}), d^4 ions (Cr^{2+} , Mn^{3+}) d^6 ions (Fe^{2+}), and d^9 ions (Cu^{2+}) splits into two, and these ions will



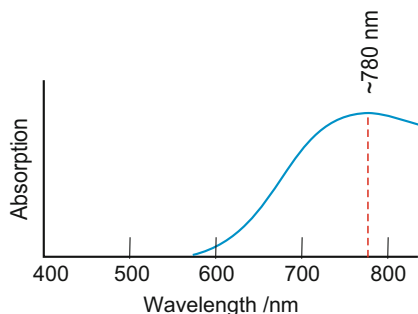
Transition Metal Ion Colors, Fig. 6 Splitting of ground state free ion terms in crystal fields of octahedral and tetrahedral symmetry. Transitions that give rise to color

shown as *arrows*. (Subscripts g should be omitted for tetrahedral coordination)

Transition Metal Ion Colors, Table 4 Octahedral crystal field absorption peaks

Ion	Crystal field terms ^a	Absorption peaks	Position/nm	Comments
Ti ³⁺ d ¹	² T _{2g} ² E _g	² E _g ← ² T _{2g}	493	[Ti(H ₂ O) ₆] ³⁺ ; violet
V ⁴⁺ d ¹	² T _{2g} ² E _g	² E _g ← ² T _{2g}	769	VO[(H ₂ O) ₅] ²⁺ ; blue
V ³⁺ d ²	³ T _{1g} ³ T _{2g} ³ A _{2g}	³ T _{2g} ← ³ T _{1g}	575	In Al ₂ O ₃ ; blue
			397	³ T _{1g} ← ³ T _{1g} from ³ P
		³ A _{2g} ← ³ T _{1g}	294	
Cr ³⁺ d ³	⁴ A _{2g} ⁴ T _{2g} ⁴ T _{1g}	⁴ T _{2g} ← ⁴ A _{2g}	556/575	In Al ₂ O ₃ ; [Cr(H ₂ O) ₆] ³⁺ ; red, green
		⁴ T _{1g} ← ⁴ A _{2g}	400/408	
Cr ²⁺ d ⁴	⁵ E _g ⁵ T _{2g}	⁵ T _{2g} ← ⁵ E _g	714	[Cr(H ₂ O) ₆] ³⁺ ; blue
Mn ³⁺ d ⁴	⁵ E _g ⁵ T _{2g}	⁵ T _{2g} ← ⁵ E _g	461/476	MnF ₆ ; [Mn(H ₂ O) ₆] ³⁺ ; red violet
Mn ²⁺ Fe ³⁺ d ⁵	⁶ A _{1g}	None		
Fe ²⁺ d ⁶	⁵ T _{2g} ⁵ E _g	⁵ E _g ← ⁵ T _{2g}	1000	In [Fe(H ₂ O) ₆] ²⁺ , pale green
Co ²⁺ d ⁷	⁴ T _{1g} ⁴ T _{2g} ⁴ A _{2g}	⁴ T _{2g} ← ⁴ T _{1g}	1176	In [Co(H ₂ O) ₆] ²⁺ ; pink, red
			500	⁴ T _{1g} ← ⁴ T _{1g} from ⁴ P
		⁴ A _{2g} ← ⁴ T _{1g}	560	
Ni ²⁺ d ⁸	³ A _{2g} ³ T _{2g} ³ T _{1g}	³ T _{2g} ← ³ A _{2g}	1176/935	In [Ni(H ₂ O) ₆] ²⁺ /[Ni(NH ₃) ₆] ²⁺ ; green
		³ T _{1g} ← ³ A _{2g}	741/571	
			429/355	
Cu ²⁺ d ⁹	² E _g ² T _{2g}	² T _{2g} ← ² E _g	780	In [Cu(H ₂ O) ₆] ²⁺ ; blue

^aDerived from ground state term for a free ion in an octahedral field, listed in ascending order

**Transition Metal Ion Colors, Fig. 7** The absorption spectrum of Cu²⁺(H₂O)₆

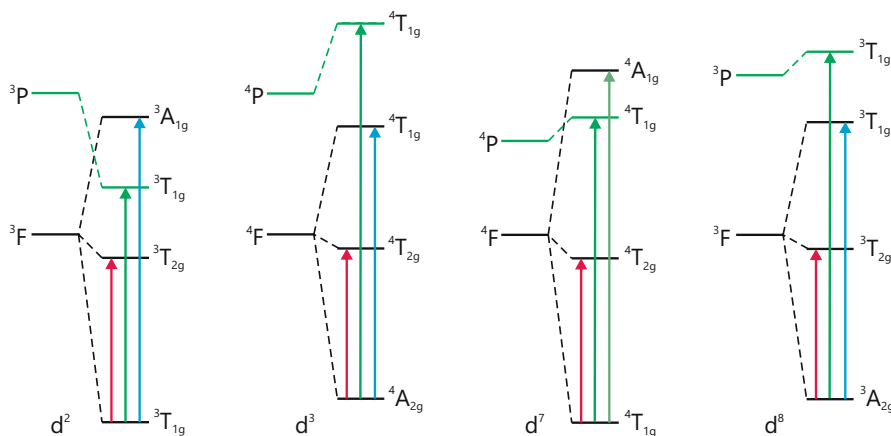
show one absorption band, corresponding to a transition between these two energy levels. If all or part of this band is in the visible, compounds containing these cations will show color (Table 4). For example, the spectrum of (approximately) octahedrally coordinated Cu²⁺ in Cu²⁺(H₂O)₆ has a single absorption band in the infrared, with a peak at approximately 780 nm (Fig. 7), but this extends into the visible, resulting in absorption of red and orange, with a result that the solution takes on the characteristic Cu²⁺ blue-green color.

The situation with the ions d² (V³⁺), d³ (Cr³⁺), d⁷ (Co²⁺), and d⁸ (Ni²⁺) is not so simple. Figure 6 indicates that these should show two transitions,

but the experimental data (Table 4) makes it clear that three transitions are observed. As transitions giving rise to color are between states of the same multiplicity, for these ions it is necessary to take into account the existence of other terms with the same multiplicity as the ground state. The ions d² (V³⁺), d³ (Cr³⁺), d⁷ (Co²⁺), and d⁸ (Ni²⁺) have two free ion terms of the same multiplicity, ³F ³P for d² and d⁸ and ⁴F ⁴P for d³ and d⁷ (Table 2). In the crystal field, the higher energy ³P or ⁴P states give rise to ³T and ⁴T states (Fig. 8). Spin-allowed transitions to these additional energy levels give rise to three absorption bands in the case of V³⁺, Co²⁺, and Ni²⁺ (Table 4). In Cr³⁺ ions the transition to the ⁴T state is at a high energy and is not recorded in the visible spectrum. These absorption peaks will give rise to colored compounds, but the shade will depend upon the position and relative intensity of the bands. This can be quite variable, as described for Cr³⁺ in the following section.

The Effect of Crystal Field Strength and Other Factors

Although crystal field splitting gives an accurate picture of the number of absorption peaks observed, the actual position of these peaks



Transition Metal Ion Colors, Fig. 8 Splitting of ground state and higher free ion terms with the same multiplicity in a crystal field of octahedral symmetry, schematic. Transitions that give rise to color shown as arrows

depends upon the crystal chemistry of the surroundings, which alters the effective strength of the crystal field. Moreover, the absorption peaks are frequently broad or show shoulders, indicating that the simple description given above is in need of refinement. Here just one example will be given. Octahedrally coordinated Cr^{3+} ions can induce a variety of colors including ruby red, emerald green, and violet. In all cases the color arises from transitions between the $^4A_{2g}$ ground state and the higher $^4T_{2g}$ and $^4T_{1g}$ levels. Ruby consists of colorless corundum (Al_2O_3) crystals with $\sim 0.1\%$ of Cr^{3+} ions occupying octahedral Al^{3+} sites. In this material the absorption peaks are:

$$^4T_{2g} \leftarrow ^4A_{2g}; \text{maximum } 556 \text{ nm; green-yellow}$$

$$^4T_{1g} \leftarrow ^4A_{2g}; \text{maximum } 400 \text{ nm; violet.}$$

The absorption at 556 nm removes green-yellow and the absorption at 400 nm removes violet. Between the absorption curves, there is a relatively small transmission window at approximately 486 nm, and a red transmission window is present at wavelengths greater than 650 nm. This means that the color transmitted by the ruby will be red with something of a blue undertone.

Emerald consists of the colorless mineral beryl with Cr^{3+} impurity ions occupying octahedral Al^{3+} sites. In beryl the octahedra surrounding the Cr^{3+} ions are slightly larger than in corundum, and

so the crystal field experienced by the Cr^{3+} in emerald is weaker than in ruby, causing a shift in the $^4T_{1g}$ and $^4T_{2g}$ levels toward the ground state with the result that the transitions move slightly toward the lower end of the spectrum:

$$^4T_{2g} \leftarrow ^4A_{2g}; \text{absorption maximum } 650 \text{ nm; orange-red}$$

$$^4T_{1g} \leftarrow ^4A_{2g}; \text{absorption maximum } 450 \text{ nm; blue.}$$

The band that absorbs green-yellow in ruby now absorbs orange-red in emerald, and the violet-absorbing band in ruby now absorbs blue. Between the absorption curves, there is a blue-green transmission window at 500 nm. The result of these small shifts is to transform the ruby red into emerald green.

In crystalline chrome alum and water solution, the Cr^{3+} ions are at the center of an octahedron of O^{2-} ions in the complex $[\text{Cr}(\text{H}_2\text{O})_6]^{3+}$. The crystal field now gives rise to the absorption bands:

$$^4T_{2g} \leftarrow ^4A_{2g}; \text{absorption maximum } 575 \text{ nm; yellow}$$

$$^4T_{1g} \leftarrow ^4A_{2g}; \text{absorption maximum } 408 \text{ nm; violet.}$$

These bands are positioned so as to emphasize purple violet, the dominant tone of chrome alum.

Although the overall color of transition metal ions in solids is explained by crystal field effects, the eye can detect subtle variations that require more complex explanations. Again ruby serves as an example. In this gemstone, the octahedra that are occupied by the Cr^{3+} ions are distorted, causing the ${}^4\text{T}_{2\text{g}}$ and ${}^4\text{T}_{1\text{g}}$ levels to split due to a change of local symmetry. The separation of the new levels is small but does give a color change that adds to the value of the stones and also produces polarization effects. In addition, an energy level ${}^2\text{E}_{\text{g}}$ (derived from splitting of the free ion ${}^2\text{G}$ term) falls between the ground state and the ${}^4\text{T}_{2\text{g}}$ level. This level is also split in the distorted octahedral sites in these crystals. Although direct excitation to and from the ground state to the ${}^2\text{E}_{\text{g}}$ levels is forbidden by the multiplicity selection rule, it is populated indirectly, and that gives rise to two closely spaced emission lines R_1 at 693.5 nm and R_2 at 692.3 nm in the ruby spectrum. This radiation enhances the color of the best rubies and is made to dominate light emission during ruby laser action.

Color as a Structural Probe

The absorption spectrum of a transition metal ion in a solid can give structural and chemical information that is not easily obtained via other techniques. This is because the absorption spectrum depends upon the local geometry of the site occupied and the valence state of the colored ion. Although this information is sometimes available from X-ray diffraction, in crystals where several cations with more than one valence state may statistically occupy the same site, X-ray diffraction data can be ambiguous. Moreover, the measurement of absorption spectra is diagnostic of the oxidation state of transition metal ion and can be used to infer the local conditions prevailing during formation.

For example, the spinel structure is adopted by many compounds with a formula AB_2O_4 , where A and B are medium-sized cations. In this structure the cations sit in both octahedral and tetrahedral sites. The absorption spectra of transition metal ions are quite different for these geometries, and so the site occupancy can be easily and unambiguously determined. The spinel NiAl_2O_4 is a case in point. The absorption

spectrum of this material reveals that the Ni^{2+} ions are found in both positions. In the related spinel NiGa_2O_4 , the Ni^{2+} ions exclusively occupy octahedral sites.

The problem of site occupancy is more acute in amorphous or disordered materials such as glasses because these structures cannot readily be determined by X-ray crystallography. However, it is often possible to incorporate a small amount of a transition metal into the structure as a probe of local geometry. For example, silicate glasses containing a small quantity of Co^{2+} are colored blue, typical of tetrahedral coordination, indicating that these ions replace Si^{4+} in the glass network. Similarly, small amounts of Mn^{2+} , Co^{2+} , and Fe^{2+} incorporated into a ZnCl_2 glass show that in each case, the absorption spectrum corresponds to that expected from tetrahedrally coordinated ions and indicates that the structure is formed from a random network of linked ZnCl_4 tetrahedral units.

Cross-References

- [Inorganic, Hybrid, and Functional Pigments](#)
- [Lanthanoid Ion Color](#)
- [Pigment, Ceramic](#)

References

1. Smith, D.W.: Ligand field theory and spectra. In: King, R.B. (ed.) *Wiley Encyclopedia of Inorganic Chemistry*, 2nd edn. Wiley, Chichester (2005)
2. Tilley, R.J.D.: Chapter 7: Colour from atoms and ions. In: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: Chapter 4. In: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Schriver, D.F., Atkins, P.W., Langford, C.H.: Chapters 6, 14. In: *Inorganic Chemistry*, 2nd edn. Oxford University Press, Oxford (1994)
5. Newman, D.J., Ng, B. (eds.): *Crystal Field Handbook*. Cambridge University Press, Cambridge (2001)
6. Henderson, B., Bartram, R.H.: *Crystal-Field Engineering of Solid State Laser Materials*. Cambridge University Press, Cambridge (2000)
7. Schäfer, H.L., Gliemann, G.: *Basic Principles of Ligand Field Theory*. Wiley, New York (1969)
8. Figgis, B.N., Hitchman, M.A.: *Ligand Field Theory and Its Applications*. zWiley-VCH, New York (2000)

Translation of Color Language

Mari Uusküla

School of Humanities, Tallinn University, Tallinn, Estonia

Synonyms

Color language translation; Color term translation; Color word translation; Glossing; Translation of color terms; Translation of color words

Definition

An action in which a color term or color phrase is converted into the meaning of color sense mutually intelligible to the speakers of two different languages, between different historical stages of one or more languages or, in some rare cases, to the speakers of one and the same native language.

Translation of Color Terms Between Languages

As an interdisciplinary subject, the translation of color language has been addressed in linguistics (cognitive linguistics, psycholinguistics, ethnolinguistics, semantics, and historical linguistics), translation studies, semiotics, social anthropology, philosophy, and some other disciplines. Besides translating between languages, the translation of color language can also refer to translating between different historical phases in one language or between a pair of languages. Translation of color terms can be considered in literal language or figurative language. In some rare cases, one can also allude to intra-linguistic translation, that is, a translation of color language within one language when the speakers of the same language disagree on the borders of color categories, naming the same color inconsistently.

Translation is a meticulous action to be carried out by a person who translates from one language

into another, and therefore, many aspects should be considered: the linguistic and social context, cultural customs and practices. The most common type of translation of color language is indisputably a translation performed in two different languages, either genetically related or not, e.g., English *green* can be translated as *verde* in Italian. In translation studies, the language from which one is translating is also called/labelled the *source language* (English in the previous example) and the language into which one is translating is labelled as the *target language* (Italian in the previous example). Besides being a linguistic concept, color is a cultural concept; therefore, the task of translating a color term into another language frequently poses a linguistic and cultural dilemma to a translator [1].

According to the universalist theory or the so-called Berlin and Kay thesis, color language is relatively similar in all languages as all languages possess a certain number of basic color terms from minimally 2 to the maximum of 12 [2]. Yet, the weak relativist theory maintains that there are divergences in both the number of basic color terms and the focals or best examples to which these color terms refer to [3, 6]. Recent studies have suggested a possible augmentation of the number of basic color terms, for example, in English there are two new candidates, *lilac* and *turquoise*, enlarging the number of basic color terms in English to 14 [4].

The question of the translation of color terms was already posed in the original study by Berlin and Kay, but from a different perspective. Berlin and Kay concluded from their examination of 98 languages that the words of color “translate too easily among various pairs of unrelated languages,” which supported the foundation of linguistic universals in the society where the “extreme linguistic relativity” prevailed [5]. However, a significant body of research has challenged the claim of effortless translation between languages, indicating several inconsistencies in mapping color terms into the system of another language, especially between the languages which carve up the color space differently, e.g., Berinmo with its 5 basic color terms and English with at least 11 [3, 6].

Moreover, English has only one color term to designate the blue area of the color spectrum, and therefore, one basic color term for blue. Russian or Italian, on the contrary, both have two to three basic color terms, covering the same area of the color spectrum as the English blue. This complicates the task in which one has to translate the English *blue* into Russian or Italian. The English *blue* might correspond to either *sinij* or *goluboj* in Russian and *blu*, *azzurro*, or *celest*e in Italian [7, 8]. In the classic example, Louis Hjelmslev compares the alignment of color terms in English and Welsh. Namely, the Welsh term *glas* covers parts of *green*, *blue*, and *gray* in English [9]. The Welsh term *gwyrd*d can only be used for parts of green, not the whole green spectral area in English. Therefore, English *green* can either convert into Welsh as *glas* or *gwyrd*d. Translators must be aware of these cases and bring the right terms in line to perform a correct translation. Ardener [10] elucidates the classic example first published by Hjelmslev of how the Welsh speakers divide color categories by comparing Standard Welsh and Modern Colloquial Welsh with English and Igbo, as all these languages carve up the color space differently – five English color categories are translated into four color categories in Standard Welsh and only into two categories in Igbo (Table 1 adapted from Ardener). The key here probably lies in the aspect that hue-based categories (English, Modern Colloquial Welsh) are translated into brightness-based categories (Standard Welsh and Igbo) [6, 11].

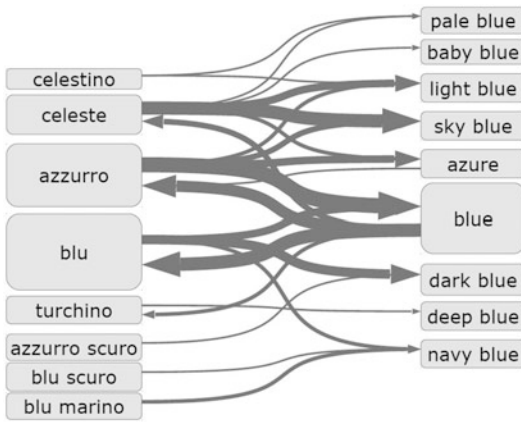
As important as the translation of basic color terms between languages may seem, the translation of color language does not only include the

translation of basic color terms but also encompasses the translation of nonbasic color terms. Basic color terms constitute only a small part of color language, and therefore, translatable words often belong to the category of nonbasic color terms [13, 14]. In particular, translation of nonbasic color terms and expressions might pose a more challenging task than finding appropriate equivalent to basic color terms in a target language. This should be taken into account in performing translation from one language into another, especially if the number of basic color terms in the given language pair diverges. In order to offer adjacent translation of color terms in target and source languages, one should ideally research color terms in both languages, finding out about their semantics, lexical and morphological potentials, and the context in which color terms usually appear [12–15].

One possibility to perform translation is to look up the meaning of a color word in another language in a dictionary. However, dictionary entries might not include extensive information on all nuances of color terms and thus, the usage of dictionaries in translation of color terms should be considered with care. Tavast and Uusküla used probabilistic weighting model to account for the better translation options between Italian and Estonian and Spanish and Estonian, also retrieving information from color-naming studies, instead of relying solely to dictionary entries [15]. The problem with dictionary entries is often twofold. First, dictionary definitions might result in inconsistencies and inaccuracies, because they intend to use the basic color terms in order to point to certain objects which might be better described with nonbasic color terms, e.g., “red is the color of blood or fire” where none of these objects actually refer to the red hue but are used metonymically. Often equivalents follow orthographic similarities across languages or through homography, neglecting the actual color sense. Second, it is usually impossible to include color samples as links to certain definitions in dictionaries, especially in printed versions of dictionaries. Thus, color term definitions might remain vague or inconsistent [13]. On the other hand, metonymical expressions such as *raspberry red* or *tulip red*

Translation of Color Language, Table 1 Welsh color terms and their possible alignment to English and Igbo. (Adapted from Ref. [10])

English	Standard Welsh	Modern Colloquial Welsh	Igbo
Green	gwyrd	gwyrd	ahehe ndu
Blue	glas	glas	
Gray	llwyd	llwyd	ojii
Brown	du	brown	
Black		du	



Translation of Color Language, Fig. 1 Color terms of blue area of the color spectrum in Italian and English as represented in dictionaries [15]. The edges of the graph represent the number of dictionaries in which a certain relationship between the two words was found: the wider the edge, the more this translation version was found in dictionaries

should translate well between the languages, provided that two languages under examination have the same species or raspberries or tulips. A better approach to interpret color terms in two different languages may be to include concepts (colors themselves if possible) into the scheme. Figure 1 represents color translations performed with the help of dictionary entries.

Comparative or contrastive studies between different language pairs provide more substantial information about the meaning of color terms than dictionary entries ever can as they also study the semantic field, lexical and morphological potentials, and context-related information of certain color terms. Philip has comparatively studied the meaning of color terms in Italian and English [16], Jones has published a substantial overview of German color terms with all aspects of their meaning, usage, and historical information which surely facilitates the translation of German color terms into any other language [17].

Kim and colleagues conducted a color-naming experiment using color samples in order to train the computer to provide better translation equivalents in 14 languages instead of widely used online translation tools such as Google translate [18]. They argue that their algorithms are able to

take into account perceptual similarity of the colors matching with color terms in the source and target language. Similarly to Tavast and Uusküla, Kim and colleagues also hint to a possibility to use visualization tools in the future.

Translation of Color Terms Between Different Stages of Language(s)

Translation of color language can also refer to alignment of color terms between different eras or different historical stages of one or more than one languages. This can take place, as in synchronic translation, between two different languages (e.g., Latin and Italian) or across different phases of one language (e.g., Old English and Modern English) or even be restricted to one phase of a certain language (Modern English). In two first cases (Latin to Italian and Old English to Modern English), we find more references to *etymology* than translation in the literature. Translation between different eras can be an extremely difficult task for translators as they need to have substantial knowledge of the context of the historical language. The most difficult are the cases in which there are no written documents of a historical stage of a language from which one is “translating” color language into the present sense. The researcher who is “translating” the color terms in the past to the current variety of a language will benefit from having a dictionary which provides large number of examples in historical texts with the dates of the earliest and the latest occurrences. It is also important to study the lexical, morphological, and syntactic patterns and changes in the past and current versions of a language [12]. According to Biggam, it always depends whether the color system is based on hue, lightness/darkness, or other aspects of appearance. Specific problems can be targeted in the later stages of languages, because we possess more information about the language and the culture if we have more written sources. Table 2 presents all possible Latin equivalents to Old English *græg* (Modern English *gray*) and their possible meaning is Latin [12, 19]. There are altogether eight possible translations to Old English *græg* in

Translation of Color Language, Table 2 Latin translations of Old English *græg* ([19]: 36)

Old English	Modern English	Latin translation	The meaning of Latin translations
<i>græg</i>	<i>gray</i>	<i>cinereus</i>	Ash-gray, ashen
		<i>croceus</i>	Saffron-colored, yellow (or ruddy)
		<i>cycnēus</i>	Swan-like (in whiteness)
		<i>elbus</i>	Dull yellow, gray or brown
		<i>ferrūgineus</i>	Color of rust, purple or black; iron-colored, gray
		<i>ferrūgō</i>	Color of rust, purple or black; iron-colored, gray
		<i>fuscus</i>	Dark-colored; (of person) dark-skinned; gray; obfuscated, opaque
		<i>glaucus</i>	Gray, grayish-blue, silvery-gray; shiny, sparkling, brilliant; grayish-yellow, yellow or orange

Latin. Apparently, the meaning of Old English *græg* has been wider from the color sense of *gray* in Modern English, containing not only grays, but expanding the color sense of the color term considerably.

“Translation” of Color Terms Within One Language

The Berlin and Kay tradition predominantly considers color language without variation of its speakers. However, multiple studies have referred to differences in color lexicon of males and females, mono-, bi-, and multilingual speakers, older and younger speakers of a language, as well as native speakers living in the cities or countryside [12]. Those variations between the speakers of the same native language, having different social, educational, and economic backgrounds, can sometimes be indicated to as intra-linguistic translation or translation of color terms within one language and within one community. Intra-linguistic translation can also refer to inconsistencies, even disagreement of color categories within one linguistic community. In most European languages, there may be vast inconsistencies on whether the same area of color spectrum could be called green or blue – that depends on the color naming alone. Sutrop points to a color term *emerald* which might be either green or blue for the speakers of the same native language [14]. Moreover, Estonian *sinakasroheline* “bluish-green” and *rohekassinine* “greenish-blue” can

actually refer to the same color in the color space, despite of belonging to the “blues” or to the “greens.” Speakers of different dialectal varieties of the same language can also tend to speak differently about colors [11]. This claim supports the fact that everyday communication is based on the frequent occurrence of specific color terms that are applied for some daily concepts which might not be present in another culture or community (e.g., the colors of national flags, specific food or beverages, fruits or greenery, etc.).

Translation of Colors in Figurative Language

As language is often viewed in relation to cognitive processes and their speakers, the translation of color language has been studied both in literal language (as taken into account beforehand) and in figurative language. Figurative language refers to idioms, metaphors, metonyms, or larger units of text, such as phrases, and is also used in humor and puns. Often it is difficult to make a difference between diverse types of figurative language, sometimes all these types are referred to as color metaphors or color idioms.

One characteristic of color terms in figurative language is that color terms are no longer designating a perceptual hue, e.g., “red” in expression “he was caught red-handed” where “red-handed” does not refer to red hue per se [13]. Not all color terms participate in figurative expressions in the same proportion and some color terms are even

rare to be found in such expressions, e.g., color term *orange*, while other color terms are used in figurative expressions in large proportions, e.g., *black*, *white*, and *red*, despite the language. However, different languages possess different color metaphors and translating them from source language into target language might cause translation dilemmas and one should always rely on the context of color terms used figuratively.

Four types of translation strategies can be discerned in translation of figurative color expressions [20]. First, reproducing the color metaphor with a coequal color metaphor in the target language (e.g., *rose-colored glasses* remains *rose-colored glasses* in a target language); second, replacing the color metaphor without a color word, but preserving a metaphorical expression in the target language (e.g., English color metaphor *blue in the face* could be translated into various target languages using other color words (e.g., *green*, *red*) or replacing a color word with random metaphorical expression); third, converting the color metaphor to a paraphrase (e.g., English metaphors *purple prose* or *yellow-bellied person* seem to be language specific and often need to be paraphrased into other languages); and fourth, omitting the color metaphor if it is redundant or untranslatable. Each language has color metaphors that are shared by different languages and widely understood in diverse cultures, but there are also language- and culture-specific metaphors which do not have equal counterparts in other languages and cultures. It seems that the more novel the metaphor, the less it is shared widely between different languages and cultures. In any case, metaphor translation should always be carried out meticulously, taking into account many linguistic and cultural factors.

References

1. Eco, U.: *Experiences in Translation*. University of Toronto Press, Toronto (2001)
2. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
3. Roberson, D., Davies, I.R.L., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* **129**(3), 369–398 (2000)
4. Mylonas, D., MacDonald, L.: Augmenting basic colour terms in English. *Color Res. Appl.* **41**, 32 (2016)
5. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969/1991)
6. Saunders, B.: *The Invention of Basic Colour Terms*. Utrecht (1992)
7. Paramei, G.: Singing the Russian blues: an argument for culturally basic color terms. *J. Cross-Cult. Res.* **39**(10), 10–38 (2005)
8. Grossmann, M., D'Achille, P.: Italian color terms in the BLUE area: Synchrony and Diachrony. In: Silvestre, J.P., Cardeira, E., Villalva, A. (eds.) *Colour and Colour Naming: Crosslinguistic Approaches*, pp. 21–50. Centro de Linguística da Universidade de Lisboa – Universidade de Aveiro (2016)
9. Hjelmslev, L.: *Prolegomena to a Theory of Language*. Translated by Francis J. Whitfield. University of Wisconsin Press, Madison/London (1969)
10. Ardener, E.: *Social Anthropology*. Tavistock, London. Introductory Essay: *Social Anthropology and Language*, pp. 14–107 (1971)
11. MacLaury, R.: From brightness to hue: an explanatory model of color-category evolution [and Comments and reply]. *Curr. Anthropol.* **33**(2), 137–186 (1992)
12. Biggam, C.: *The Semantics of Colour: A Historical Approach*. Cambridge University Press, Cambridge (2012)
13. Wyler, S.: *Colour and Language: Colour Terms in English*. Gunter Narr, Tübingen (1992)
14. Sutrop, U.: Towards a semiotic theory of basic colour terms and the semiotics of Juri Lotman. In: Biggam, C.P., Hough, C.A., Kay, C.J., Simmons, D.R. (eds.) *New Directions in Colour Studies*, pp. 39–48. John Benjamins Publishing, Amsterdam/Philadelphia (2011)
15. Tavast, A., Uusküla, M.: How blue is azzurro? Representing probabilistic equivalency of colour terms in a dictionary. *Color Cult. Sci.* **4**, 43–48 (2015)
16. Philip, G.: *Colouring Meaning: Collocation and Connotation in Figurative Language*. John Benjamins Publishing, Amsterdam/Philadelphia (2011)
17. Jones, W.J.: *German Colour Terms: A Study in their Historical Evolution from Earliest Times to the Present*. John Benjamins Publishing, Amsterdam/Philadelphia (2013)
18. Kim, Y., Thayer, K., Gorsky, G.S., Heer, J.: Color names across languages: salient colors and term translation in multilingual color naming models. In: Paper presented at the 21st EG/VGTS Conference on Visualisation, University of Porto, Portugal, 3–7 June 2019 (2019)
19. Biggam, C.: *Grey in Old English: An Interdisciplinary Semantic Study*. Runetree Press, London (1998)
20. Kalda, A., Uusküla, M.: The role of context in translating colour metaphors: an experiment on English into Estonian translation. *Open Linguist.* **5**, 690–705 (2019)

Translation of Color Terms

► Translation of Color Language

Translation of Color Words

► Translation of Color Language

Translucency

► Transparency

Transparency

Osvaldo Da Pos
Department of General Psychology, University of
Padua, Padua, Italy

Synonyms

Translucency

Definition

Transparency/translucency: (a) the degree of visibility of an object through a medium; (b) the property of a material or substance by which objects may be seen through that material or substance. Transparency occurs when objects are clearly perceived through the medium, while translucency occurs when they appear hazy. Very often the term transparency is used instead of translucency.

General

The episcotister-based model is suitable to describe the perception of transparency as it

involves different areas separated by clear margins, while variations of this model are required to deal with translucency due to blurred edges which add some degree of opacity to the impression of transparency [1].

First, figural and topological conditions, not discussed here, must be met: they essentially consist of figural unity of the transparent layer, continuity of the boundary line, and adequate stratification. Another important factor which strengthens the figural organization is movement, either of the background or of the transparent medium; this in turn makes transparency more evident. Transparency can also disambiguate objects (plaids) moving in different directions which otherwise would appear as a structure moving in another direction. Devices which induce strong depth perception, like natural or simulated stereopsis, also increase the appearance of transparency.

Transparency involves the distinct perception of two different colors, that of the background in the rear and that of the transparent layer in front where they are superimposed.

This part of the definition is very restrictive as there are cases in which one sees objects behind and through another object, but the colors perceived in the overlapping area amazingly differ from those in plain view. In these cases, no color constancy is achieved, and no model seems presently applicable.

History of the Model

Models are mathematical interpretations of physical phenomena, with characteristics analogous to those to be described, whose explanatory structure is heuristically applied to not yet fully understood phenomena (for instance, the atomic planetary model by Rutherford, the hydraulic analogy of electricity, and so on). A different definition of models includes just the mathematical formulation of relevant relationships which describe the phenomenon.

The episcotister model of transparency derives from experiences based on rotating wheels, used to produce different colors by mixing two or more

sources of radiation. Most times the disk is divided into two or more differently painted parts. The disk is rotated at a fusion speed where the human visual system cannot distinguish the different parts and generates the perception of only one color. This method of mixing radiation was made famous by Maxwell from whom it took the name. Sometimes, instead of a solid disk, only two symmetrical sectors are spun in order to mix rays from the background with those from the solid sectors. Although the laws governing these kinds of mixtures are well known, the actual perception does not always conform with the prediction of seeing one resulting color, as sometimes two colors are seen in the overlapping area. This outcome happens when some figural conditions are fulfilled and the figure of a transparent disk is visible in front of a background completely or partially covered and stratified in depth. This experience led to theoretical and experimental analysis of the psychological impression of transparency.

Field of Applications

The concept of transparency has been applied to food and drink analysis, skin and cosmetics, medicine, dentistry, architecture, textile manufacturing, art, graphics, design, fashion, painting, printing, and others.

Overview

Metelli's Episcotister Model of Perceptual Transparency

Metelli's model [2] consists of a wheel, rotating at fusion speed in front of a background, divided into two parts: an empty sector (totally transmitting area) and a solid sector (reflecting area). It is a kind of special filter, like a riddled surface, which transmits and reflects light at the same time.

Through the open part of the episcotister (or through the holes of the riddled surface), light coming from the object behind it can reach the observer's eye without being spectrally modified. The size of the open sector (or of the holes in

the riddled surface) determines how much light is transmitted to the observer: from a maximum when the whole disk is completely open to zero when the whole disk is solid. Therefore, the episcotister and the riddled surface behave like a neutral filter by decreasing the intensity of the radiation passing through them by a certain amount without changing its spectral distribution.

In turn, the solid part of the episcotister (or the solid part of the riddled surface) reflects radiation to the observer, and this is mixed with the radiation from the back. As both the open and the filled sectors of the episcotister (or the holes and the solid part of the riddled surface) together cannot be larger or smaller than the whole disk (or surface), their proportion can vary from 0 to 1: if α is the proportion of the open sector to the whole disk, $(1-\alpha)$ is the proportion of the solid sector (or of the solid riddled surface). As one increases, the other decreases accordingly: for this reason, the mixture of the two lights, that coming from the background and that coming from the surface of the episcotister (or riddled surface), is a special kind of mixture called partitive. If b is the vector describing the radiation coming from the background not modified by a filter and t is the vector describing the radiation coming from the solid surface of the filled episcotister, the presence in the rotating disk of an open sector of size α determines the partitive mixture described by the following equation:

$$q = \alpha \times b + (1 - \alpha) \times t \quad (1)$$

where q is the vector describing the mixture of the two radiations (b and t) arriving at the observer's eye.

This equation perceptually means that the back color b is visible in the color q proportionally to α and the color corresponding to t is visible in q proportionally to $(1-\alpha)$. The reduced visibility of the two colors due to the transparency factor means that they are not perceived as completely in plain sight, but one appears transparent and the other seen by transparency. Both, the radiation coming from the back opaque objects and the front transparent object, arrive at the observer reduced in intensity but unchanged in their

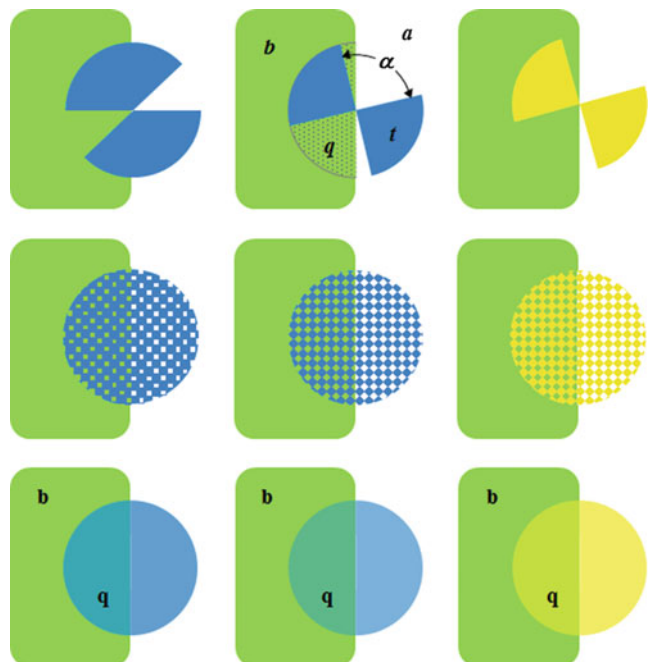
spectral composition. Therefore α indicates how much color of the background is visible in the overlapping area: this dimension corresponds to the degree of perceived transparency of the front object and, according to Metelli, is a logarithmic function of it, while α is a neutral (achromatic) multiplicative factor. The additive factor t , which can be interpreted as either the reflectance characteristic of the object or the light (luminance) reflected from it, corresponds to the color of the filter and can assume any value from 0 to 1 (black and white, respectively).

If the episcotister rotates over a uniform background, it appears like an opaque disk, and the same happens when physically transparent sheets are superimposed over a uniform surface. This effect shows that physical transparency is not a sufficient condition to perceive transparency. Another, particular, case of transparency is obtained when a transparent sheet completely covers a background figure without sharing its' margins: this effect has been little studied [3]. The case of riddled surfaces can be extended, by analogy, to textured surfaces, in which the density of the texture plays the role of color.

Equation 1 can be interpreted in two different ways, by analogy with a chemical reaction: when reading it from right to left, it means that the partitive mixture of the two colors b and t gives rise to the resulting and univocally defined color q ; but if one reads it from left to right, the equation can describe a color scission, that is, the color q can partitively split into the two components b and t . This scission, described by many researchers, would be in the model the opposite process of fusion and occurs in the perception of two colors in the same area, one in front and the other in the back. In different words, this scission would be a case of double representation in Gestalt terms. In this latter case, however, an indefinite number of b and t pairs can be obtained because α is not determined. Therefore, at least two equations are necessary to derive a specific pair of b and t , and a second background color seen through the transparent object must be present. For instance, in Fig. 1, the episcotister should not only rotate over the green background figure but also over the white background a , and in this case another partitive mixture would be described by the following equation:

Transparency,

Fig. 1 Top row: three different episcotisters over three identical backgrounds. b = vector defining the color of the background; t = vector defining the color of the solid sector of the episcotister; q = vector defining the color of the overlapping area; α = a measure of the open sector of the episcotister, which can vary from 0 to 1. Middle row: three riddled circular surfaces. The holes/total surface ratio (α) can vary from 0 to 1. Bottom row: the perceptual effects



$$p = \alpha \times a + (1 - \alpha) \times t \quad (2)$$

where p is the vector describing the mixture of the two radiations (a and t) arriving at the observer's eye.

The determination of α and t , characteristics not present in the stimulation, can be univocally found, starting from the colors of the display in the bottom row of Fig. 1 according to the following equations:

$$\alpha = (p - q)/(a - b) \quad (3)$$

$$t = ((aq) - (bp))/(a + q - p - b) \quad (4)$$

Once the a (the lightest area which is perceived in the background) and b background colors and t , the color of the transparent virtual object, have been chosen and the degree α of transparency is selected, the regions of the mosaic of Fig. 2 are filled with the colors determined according to Eqs. 1 and 2.

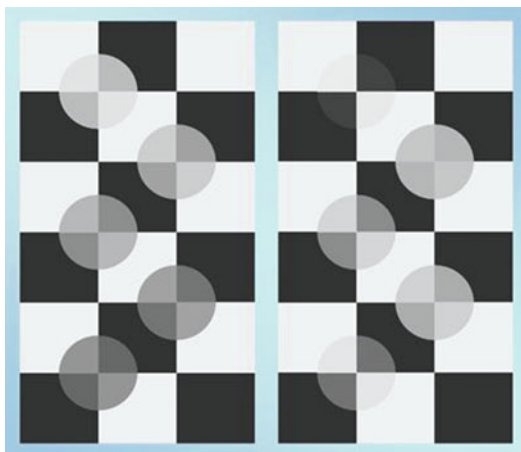
The colors of the transparent disks and their degree of transparency can be computed according to Eqs. 3 and 4 from the colors of Fig. 2. The formal independence of α and t is a feature of the model. The square wave margins between colors in the overlapping areas are characteristics of the transparency effect, while blurred margins denote translucency [3, 4]. The

alpha blending procedure in computer graphics is based on this model [5].

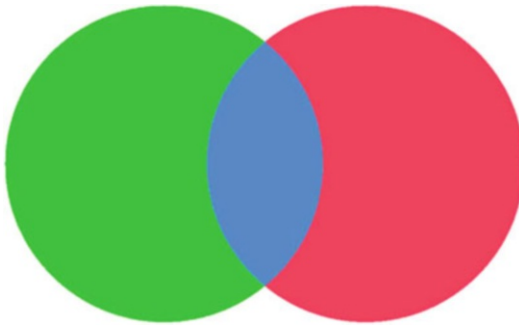
The model of the episcotister applied to transparency phenomena entrains also some constraints and consequences. The most important is that α [5] cannot be higher than 1 or smaller than 0 (an open sector beyond 360° or below 0° is inconceivable). Thus, by assuming a is the vector of the lightest color in the background and b the other background color (both perceived as such), the difference $p - q$ must be smaller than the difference $a - b$: the contrast (here measured as a color difference) between the colors covered by the transparent object is always lower than the contrast between the two background colors in plain view. In other words, the transparent object reduces the color contrast in the overlapping area. Secondly, as α must be positive (a negative open sector is a nonsense), the color difference between p and q must go in the same direction of the color difference between a and b . If these conditions are not met, usually transparency cannot be perceived, and the model cannot be applied, i.e., nothing can be stated about the possible transparency effect on this basis.

Transparency and Color Constancy

Transparency, like illumination, involves color constancy: first, the colors of the background perceived through the transparent medium should appear the same as those in plain view, and secondly, the color of the transparent medium should also appear the same in all its parts. The analogy with illumination is justified by the often quoted example of a transparent medium as a spectrally selective filter. Models based on simple filters have been shown not to work because the colors of the overlapping regions can be quite unexpected and sometimes color constancy does not hold at all: for instance, through a suitable red filter, a red surface may appear white and a green surface may appear black. Therefore, filter models usually include either the need of previous experience or some constancy constraints. A model based on the illumination analogy [6] considers the cone excitations as the basis for describing the



Transparency, Fig. 2 Transparent gray disks of different t color (at left) and different degrees (α) of transparency (at right) over a white and black chessboard



Transparency, Fig. 3 An example of impression of transparency without color constancy. (From Metzger [8]). When the disks are rotating, the impression is much stronger

necessary color conditions of an impression of transparency. This can occur when cone-excitation ratios between the colors in plain view ($a, b, \dots$) remain the same between the colors ($p, q, \dots$), i.e., under the transparent medium. Although color constancy is a strong requirement for perceiving transparency, it can be very weak or even null, but nevertheless a generic impression of transparency is still possible (Fig. 3). This makes the study of perceptual transparency quite difficult [7].

Transparency and Contrast

The model specifies a, b, t, p , and q by the achromatic reflectance characteristics of the surfaces (objects, materials, media) and describes the reduced contrast in the superimposition area by a reflectance difference, while further developments of the model specify colors by lightness Munsell values [9], by luminance [10], or by log luminance [11]. On the other side, there is no uniquely shared measure of contrast, and sometimes one can use the Michelson contrast [4] which seems to supply a better fitting model, suitable to explain why usually light filters seem more opaque and dark filters more transparent, although this model too is not generally applicable [12]. In any case, the filter reduces the contrast of the colors seen behind it because as they get closer, the filter gets denser (or less transparent)

[5]. Apparent contrast is further reduced by blurred edges [3].

Color contrast is relevant in determining the relative stratification of the surfaces, i.e., what is seen in front and behind, as higher contrast results in larger distances in depth. Contrast does not depend only on color difference but also on the shape of the margins: the Craik-O'Brien-Cornsweet edges (blurred versus sharpened) can invert the contrast appearance and as a consequence the relative depth of the surfaces, making transparent what was opaque and the reverse [13]. This effect can be obtained also by simple line drawings, where their contrast with the background determines the different object stratifications and the consequent transparent appearance.

A Qualitative Account of the Model in the Achromatic Domain

The possibility of perceiving transparency can be described as a function of the ordinal position of colors in the regions divided by X junctions [14, 15, 16]. If all colors present in a transparent display are achromatic, and therefore characterized by one dimension, two ordinal conditions of transparency are prescribed by the model:

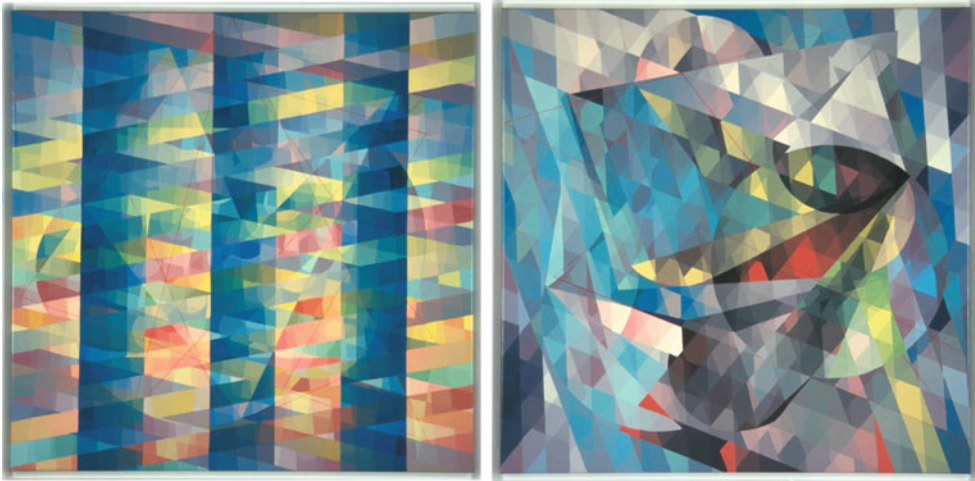
$$|a - b| > |p - q| \text{ and } p > q \text{ if } a > b \text{ (or } p < q \text{ if } a < b).$$

A better prediction of transparency is achieved with the single ordinal condition: $p \in (a, q)$ or $q \in (p, b)$ [16].

Another improvement of the model states that, instead of Eq. 3, the apparent density ψ (inverse of transparency) of the filter [17] depends on three color differences:

$$\psi = w_1 |a - p| + w_2 |b - q| + (1 - w_1 - w_2)(k - |p - q|) \quad (5)$$

where the weight coefficients w_1 and w_2 express a compromise between the three differences.



Transparency, Fig. 4 Attilio Taverna. Left: *Sentieri interrotti* 1992. Right: *Blue geometric storm* 1992. Transparency effects are obtained by painting each overlapping

area with the partitive mixture of the back and front colors. Also perceptual space organization is affected by transparency

Transparency and Pictorial Art

Since centuries, the transparency effect is used both to represent translucent or transparent objects and to enrich the color variety of a painting.

Many methods have been adopted to render transparent objects, all having in common the purpose of showing two colors, one in front and one in the back, the latter visible through the former. One method is just to put in the overlapping areas the color which would be obtained by partitive mixture of the back and the front colors (Fig. 4), following, not necessarily consciously, the episcotister model.

Another method reaches the same results by using small mosaic tesserae to be viewed at a certain distance in order to produce spatial color mixtures (Fig. 5).

Another method is the veiling procedure obtained by using semitransparent paints, as shown in Fig. 6: the back and the transparent colors are both visible where they overlap. Lastly, a final method consists of accentuating the borders between the opaque and the transparent medium, by simulating Cornsweet edges and introducing a depth displacement between opaque and transparent surfaces.

Improvements of the Model

As the colors in plain view and those in the overlapping areas, represented in a suitable color system, according to the model converge in a point corresponding to the color t , the model has been called convergence model [5]. The convergence point can lie outside the color solid, and perception of transparency can still be possible [7], although not always [18]. In this partitive model, α defines the position of the color p between the color a and the color t , and colors necessarily have to be expressed in colorimetric quantities. There are different methods, used in different theoretical models, to compute α from the colors of the image, but the result is always a value of physical nature whose relationship with the corresponding perception of transparency is not linear.

From a different point of view, the same mathematical *model* can be used without physical implications, and in this case α is a distance measure in a perceptual space. While in Metelli's model α has no direct perceptual meaning, except that of the order (more or less transparent), values of α with perceptual meaning can be derived according to traditional methods by logarithmic transformations [12]; for achromatic colors, α can also be computed by using Munsell values instead



Transparency, Fig. 5 Ravenna: Baptistery of Aryans, fifth century (detail). Picture by Incola (Wiki) (http://it.wikipedia.org/wiki/File:Soffitto_Battistero_Ariani_Ravenna.jpg#filehistory). Water and body colors seen from

some distance partitively fuse because under spatial discrimination threshold, an impression of transparency arises in the observer



Transparency, Fig. 6 Hans Memling, fifteenth century. Portrait of a young lady (Sibylla Sambetha). Musea Brugge, Sint-Janshospitaal. (Copy purchased from Lukas Art in Flanders, o.n. 1376 (2012))

of reflectance or luminance [19], since in the Munsell system equal color differences are represented by equal perceived distances. Colorimetric color difference systems (for instance OSA_UCS, CIELUV, CIELAB, CIECAM, and so on) can be used to compute a perceptual α , of course with varying scales. This latter way of computing α produces results with better perceptual agreement than the former [19], and many software programs implement it. The reason is that the color relationships entering the computations have a perceptual basis (percept-percept coupling [20]), not a physical or physiological one. The step of using perceptual measures of colors describes a radical change compared to the traditional psychophysical procedures, and rightfully it can be considered belonging to the phenomenology of colors, science inside the world of appearance. Unfortunately, the number of not well-fitting results is still unsatisfactory. A further step has been accomplished. According to current opinion, the perceived distance between the two colors a and t should not be large; otherwise transparency perception would not occur. The reason is that the color p should not be just a

Transparency, Fig. 7 In the upper part, the external areas represent, in isolation, the background in plain view and the internal areas the overlapping surfaces, whose colors have been defined in terms of the NCS. The bottom part shows the complete mosaic resulting when the separated areas are joined



colorimetric intermediate between a and t but should appear *similar* both to a and t [21]. This would be a strong condition to see both colors in the overlapping area. Large differences between a and t would not result in perceived transparency because there is no color resembling both at the same time. In these cases spatial induction can sometimes add the necessary missing shades (especially if p is a neutral color). A model has been proposed [21] in which the Natural Color System [22], based on the perceptive *similarity* of any test color to the six Hering prototypical unique colors W K Y R B G, is used to describe the relevant color relationships in transparency perception (Fig. 7).

The naming of colors in this system clearly shows similarity relationships and therefore the possibility of seeing two colors in a not-unique (i.e., perceptually mixed) color. The limits of the models are limits in color vision: no color can resembles two opponent colors at the same time in the same place, and therefore they cannot be seen in the overlapping area of a transparent object. On the other side, it is known that a weak form of transparency is possible in which the colors of the background are not clearly visible

in the overlapping areas and the transparent object does not show a uniform color [8].

Cross-References

- [Color Constancy](#)
- [Color Contrast](#)
- [Color Phenomenology](#)
- [Hering, Karl Ewald Konstantin](#)
- [Mosaics](#)

References

1. Metelli, F., da Pos, O., Cavedon, A.: Balanced and unbalanced complete and partial transparency. *Percept. Psychophys.* **38**, 354–366 (1985)
2. Metelli, F.: The perception of transparency. *Sci. Am.* **23**, 90–96 (1974)
3. Ekroll, V., Faul, F.: Transparency perception: the key to understanding simultaneous color contrast. *J. Opt. Soc. Am. A.* **30**, 342–352 (2013)
4. Singh, M., Anderson, B.L.: Photometric determinants of perceived transparency. *Vis. Res.* **46**, 879–894 (2006)
5. D'Zmura, M., Rinner, O., Gegenfurtner, K.R.: The colors seen behind transparent filters. *Perception.* **29**, 911–926 (2000)

6. Ripamonti, C., Westland, S., da Pos, O.: Conditions for perceptual transparency. *J. Electron. Imaging*. **13**, 29–35 (2004)
7. Richards, W., Koenderink, J.J., van Doorn, A.: Transparency and imaginary colors. *J. Opt. Soc. Am. A*. **26**, 1119–1128 (2009)
8. Metzger, W.: Über Durchsichtigkeits – Erscheinungen. (Vorläufige Mitteilung). *Riv. Psicol., Fascicolo Giubilare*. **49**, 187–189 (1955)
9. Beck, J.: Additive and subtractive color mixture in color transparency. *Percept. Psychophys.* **23**, 256–267 (1978)
10. Gerbino, W., Stultiens, C.I.F.H.J., Troost, J.M., de Weert, C.M.M.: Transparent layer constancy. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 3–20 (1990)
11. Vladusich, T.: A reinterpretation of transparency perception in terms of gamut relativity. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **30**, 418–426 (2013)
12. Masin, S.C., Tommasi, M., da Pos, O.: Test of the Singh-Anderson model of transparency. *Percept. Mot. Skills*. **104**, 1367–1374 (2007)
13. da Pos, O., Devigili, A., Giaggio, F., Trevisan, G.: Color contrast and stratification of transparent figures. *Jpn. Psychol. Res.* **49**, 68–78 (2007)
14. Adelson, E.H., Anandan, P.: Ordinal characteristics of transparency. In: AAAI-90 Workshop on Qualitative Vision, July 20, Boston, pp. 77–81 (1990)
15. da Pos, O., Burigana, L.: Qualitative inference rule for perceptual transparency. In: Albetazzi, L. (ed.) *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space and Appearance*, pp. 369–394. Wiley Blackwell, Chichester (2013)
16. Masin, S.C.: The luminance conditions of transparency. *Perception*. **26**, 39–50 (1997)
17. Masin, S.C., Gardonio, G.: The evaluation of the apparent density of a filter on a bicolored background. *Percept. Psychophys.* **37**, 103–108 (1985)
18. Brill, M.H.: The perception of a colored translucent sheet on a background. *Color. Res. Appl.* **19**, 34–36 (1994)
19. Beck, J., Prazdny, K.: The perception of transparency with achromatic colors. *Percept. Psychophys.* **35**(5), 407–422 (1984)
20. Epstein, W.: Percept-percept couplings. *Perception*. **11**, 75–83 (1982). <https://doi.org/10.1068/p110075>
21. da Pos, O.: *Trasparenze*. Transparency. Icone, Milano (1989)
22. NCS Colour Atlas: SIS Standardiserings Kommissionen I Sverige, 2nd ed. Stockholm (1989). SS 01 91 02

Trichromacy

► [Evolution of Primate Color Vision](#)

Tristimulus Colorimeter

► [Instrument: Colorimeter](#)

Tritanopia

David L. Bimler¹ and Galina V. Paramei²

¹School of Psychology, Massey University, Palmerston North, New Zealand

²Department of Psychology, Liverpool Hope University, Liverpool, UK

Synonyms

[Autosomal dominant inherited color vision deficiency](#); [“Blue-blindness”](#); [Blue-yellow deficiency](#); [Color blindness](#); [Color vision abnormality](#); [Dichromacy](#)

Definition

Tritanopia (from the Greek *tritos*: third + *an*: not + *opia*: a visual condition) is a congenital form of severe color deficiency (dichromacy) affecting the blue-yellow opponent color system. Tritanopes do not distinguish between colors along a specific direction in color space (tritanopic confusion lines) and are able to match all colors using two primaries (unlike trichromats who require three primaries). Tritanopia follows a dominant pedigree pattern and is linked to chromosome 7, where damage to the gene encoding the photopigment opsin results in the absence of functioning S-cones, cone cells containing the short-wavelength (S-) photopigment. Tritanopia is rare, not sex-linked, with great uncertainty as to its incidence, although some estimates place this at 1 in 13,000 or fewer in the Caucasian population. The condition can be partial rather than complete, i.e., tritanopia lies at the extreme of a continuum of tritanomaly in which some capacity remains to discriminate colors along confusion lines.

Retinal Mosaic

At the retinal level, tritanopia is an absence of functional S-cones (► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)), i.e., cone cells expressing a photopigment containing an S-opsin component, tuned for selective sensitivity to the short wavelengths of light (blue) [1, 2]. These cones comprise only about 8% of the retinal mosaic. Even in normal trichromats, they are absent from the central region of the fovea, resulting in “foveal tritanopia” affecting the perception of sufficiently small stimuli when they are fixated [3, 4].

Congenital Tritanopia: Genetic Background

Functional loss of the S-cone photoreceptor cells is caused by point mutation in the gene on chromosome 7 that codes for the S-opsin component of the cone photopigment (sensitive to shorter wavelengths) (► [Genetics of Photoreceptors, Genetics and Color Vision Deficiencies, Genes of Cone Photopigments, Genes and Cones](#)). The resulting amino acid substitution may prevent the S-opsin from functioning, or it may fold into an aberrant conformation blocking it from being transported to the ciliary segment of S-cone cells [1, 5].

The condition follows a dominant pattern of inheritance: S-cones must contain two undamaged copies of the gene to function properly [6, 7]. However, pedigrees often display variations in penetrance with incomplete tritanopia in some family members who retain some blue-yellow discrimination.

Variation in phenotype may also indicate that loss of S-cones is progressive, and incomplete, in some family members. This is supported by retinal imaging showing S-cone degeneration subsequent to formation of the retina [8].

Disease processes and exposure to environmental toxins can lead to the same end point, i.e., acquired tritanopia (► [Color Perception and Environmentally Based Impairments](#)). In

particular, the phototoxicity of prolonged exposure to intense short-wavelength light can cause S-cones to degenerate. In other forms of acquired tritanopia, the damage is localized elsewhere in the processing of the S_0 signal, such as the ganglion cells or the optic nerve [9].

Confusion Lines

Tritanopic color processing lacks the $S-(L+M)$ or S_0 signal, produced by bistratified retinal ganglion cells comparing the S-cone signal and the combined L- and M-signals (Color Vision, Opponent Theory); hence, there is no difference in appearance among colors which differ only in how much they stimulate S-cones. Examples of colors confused by tritanopes are blue and green or violet, gray, and yellow. Plotted in a chromaticity diagram (► [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)), the confused colors lie on straight “confusion lines” that converge at the tritanope copunctal point (Fig. 1).

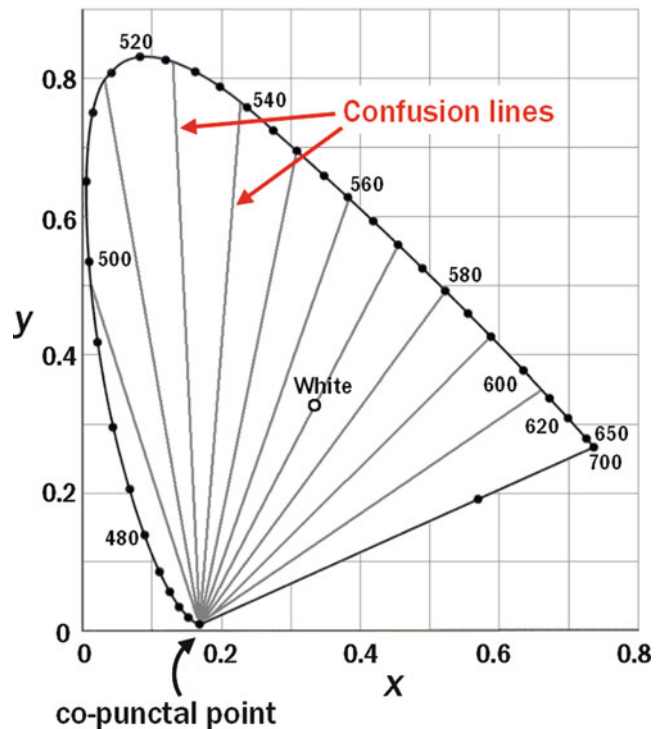
A related phenomenon is observed in edge perception, where the border between two regions of equiluminant color dissolves if the colors lie on a tritanopic confusion line (e.g., yellow and gray or green and blue). The perception of border distinctness is tritanopic, with no contribution from the S_0 signal, as a consequence of the sparseness of S-cones and chromatic aberration of the lens [10].

Wavelength Discrimination

Because S-cones are insensitive to mid- and long-wavelength lights, tritanopic hue discrimination, as a function of wavelength λ , is similar to the normal trichromat function for $\lambda > \text{ca. } 530 \text{ nm}$, with the discrimination threshold $\Delta\lambda$ reaching a minimum at about 570–580 nm near the tritanopic “neutral point” (see below). Tritanopic discrimination deteriorates for blue-green hues where the L- and M-cones are equally sensitive, with $\Delta\lambda$ peaking at ca. 470 nm and then improving at shorter wavelengths [11].

Tritanopia, Fig. 1

Confusion lines indicate colors that cannot be distinguished by a tritanope. The line going through the white point indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing S-cone fundamental. Drawn in the CIE 1931 chromaticity diagram. (Source: Benjamin, W.J. (ed.), *Borish's Clinical Refraction* (1998), Fig. 20) (Elsevier Books, Copyright Clearance Center: Licensee: David Bimler; License Number: 3532171403672)

**Luminous Efficiency**

S-cones normally contribute little or nothing to perceived luminance, but S-cone signals can make a contribution under certain conditions of adaptation and temporal flicker. Specifically, with intense long-wavelength background, S-cone stimulation exerts a negative but phase-delayed influence on luminance, as if inhibiting the combined L-and M-cone signal [12, 13].

Color Perception

A tritanope recognizes far fewer distinct colors than a trichomat (► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)). When naming spectral colors, tritanopes experience one “neutral zone” of subjective gray for light with midpoint at λ ca. 570 nm (yellow) and another for short-wavelength violets, at about $\lambda = 410\text{--}420$ nm [14]. For a tritanope, all equiluminant colors, instead of being arranged in a two-dimensional plane (► [Psychological Color Space and Color Terms](#)), form a single-

dimensional line, running from “cool” to “warm” colors between the two extremes of saturated green and saturated red. This single dimension combines the usually separate chromatic qualities of “saturation” and “hue.” Tritanopes distinguish approximately 44 steps along this line, each corresponding to a separate band of interchangeable colors [1]. This tritanopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

The subjective qualities of tritanopic experience remain unclear. Tritanopes do use the words “blue” and “yellow” when describing lights and surface colors [4, 11]. Dichromacy simulation software, which converts color scenes to show how a tritanope would see these, uses the convention of a red-green scale [15].

Diagnosis of Tritanopia

Acquired blue-yellow deficits can be symptomatic of illness or chemical exposure (► [Color Perception and Environmentally Based Impairments](#)). Thus,

standard tests of color abnormality, although targeted primarily at congenital red-green deficiency, do not ignore tritanopia completely. Pseudoisochromatic plates ([► Pseudoisochromatic Plates](#)) and arrangement/panel tests directly address a tritanope's perceived similarity of colors along the confusion axis.

The Hardy-Rand-Rittler test ([► Color Vision Testing](#)) contains pseudoisochromatic diagnostic plates to detect tritanopia and gauge its severity, while the Farnsworth F2 plate is designed specifically for that purpose (The widely used Ishihara series contain no plates targeting tritanopia). Each pattern presented in these plates is defined by a color difference that disappears for tritanopes, so that the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction.

In the arrangement tests (Farnsworth-Munsell 100-Hue test, Farnsworth D-15 test), the one-dimensionality of a tritanope's color "plane" collapses a color circle ([► Color Circle](#)) into a line, causing drastic departures from a normal observer's sequence. Errors, or cap transpositions "distances," are summed to quantify the severity of tritanopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the tritanopic angle of the confusion axis ([► Color Perception and Environmentally Based Impairments](#)).

The "gold standard" of diagnosis of tritanopia is the Moreland equation, with stimuli presented in a small viewing field (2°) in the Moreland anomaloscope [16], in which a cyan standard (480 nm with a small admixture of 580 nm) is matched by mixing indigo (436 nm) and green lights (490 nm). Decreasing discrimination along the tritanopic confusion lines increases the range of acceptable mixtures. Notably, when field size is increased beyond 1°, residual blue-yellow discrimination can be detected in some cases; at 8° only complete tritanopes accept the full range of color mixtures as a match to the cyan standard [7].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour

Test and the Colour Assessment and Diagnosis test ([► Color Vision Testing](#)), interactively quantify any loss of chromatic sensitivity along the tritanopic confusion line. For a tritanope, in the frame of the CIE (u'/v') 1976 chromaticity diagram ([► CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the tritanope copunctal point.

Cross-References

- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision Testing](#)
- [Deutanopia](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)

References

1. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Cone genes, cone pigments, color vision and colorblindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, Ch. 1, pp. 3–52. Cambridge University Press, Cambridge (1999)
2. Stockman, A., Sharpe, L.T.: Human cone spectral sensitivities and color vision deficiencies. In: Tombran-Tink, J., Barnstable, C.J. (eds.) *Visual Transduction and Non-Visual Light Perception*, Ch.14, pp. 307–327. Humana Press, Totowa (2008)
3. Williams, D.R., MacLeod, D.I.A., Hayhoe, M.M.: Foveal tritanopia. *Vis. Res.* **21**, 1341–1356 (1981)

4. Mollon, J.D.: A taxonomy of tritanopias. In: Verriest, G. (ed.) *Color Vision Deficiencies VI Documenta Ophthalmologica Proceedings Series*, vol. 33, pp. 87–101. Dr. W. Junk Publisher, The Hague (1982)
5. Neitz, J., Neitz, M.: The genetics of normal and defective color vision. *Vis. Res.* **51**, 633–651 (2011)
6. Wright, W.D.: The characteristics of tritanopia. *J. Opt. Soc. Am.* **42**, 509–521 (1952)
7. Pokorny, J., Smith, V.C., Went, L.N.: Color matching in autosomal dominant tritan defect. *J. Opt. Soc. Am.* **71**, 1327–1334 (1981)
8. Baraas, R.C., Carroll, J., Gunther, K., Chung, M., Williams, D.R., Foster, D.H., Neitz, M.: Adaptive optics retinal imaging reveals S-cone dystrophy in tritan color-vision deficiency. *J. Opt. Soc. Am. A.* **24**, 1438–1447 (2007)
9. Simunovic, M.P.: Acquired color vision deficiency. *Surv. Ophthalmol.* **61**, 132–155 (2016)
10. Tansley, B.W., Boynton, R.M.: A line, not a space, represents visual distinctness of borders formed by different colors. *Science*. **191**, 954–957 (1976)
11. Alpern, M., Kitahara, K., Krantz, D.H.: Classical tritanopia. *J. Physiol.* **335**, 655–681 (1983)
12. Stockman, A., MacLeod, D.I.A., DePriest, D.D.: The temporal properties of the human short-wave photoreceptors and their associated pathways. *Vis. Res.* **31**, 189–208 (1991)
13. Ripamonti, C., Woo, W.L., Crowther, E., Stockman, A.: The S-cone contribution to luminance depends on the M- and L-cone adaptation levels: silent surrounds? *J. Vis.* **9**(3), 10 (2009). <https://doi.org/10.1167/9.3.10>
14. Smith, D.M.: Color naming and hue discrimination in congenital tritanopia and tritanomaly. *Vis. Res.* **13**, 209–218 (1973)
15. Brettel, H., Viénot, F., Mollon, J.D.: Computerized simulation of color appearance for dichromats. *J. Opt. Soc. Am. A.* **14**, 2647–2655 (1997)
16. Moreland, J.D., Young, W.B.: A new anomaloscope employing interference filters. *Mod. Probl. Ophthalmol.* **13**, 47–55 (1974)

Tubular and Compact Fluorescent Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
Netherlands

Synonyms

Low-pressure mercury gas discharge lamps

Definition

A lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a low-pressure mercury vapor that is contained in a transparent tube whose inside is coated with fluorescent powder that converts the ultraviolet part of the emitted radiation from the discharge in visible light.

Types of Tubular Fluorescent Lamps

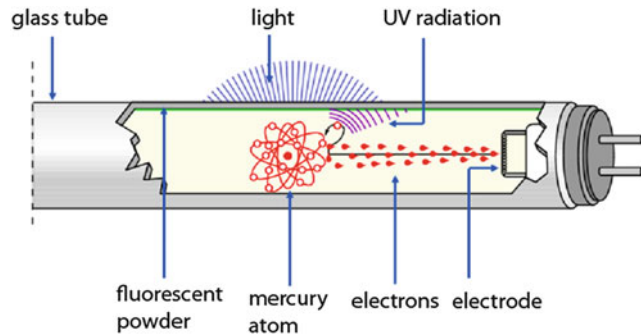
Fluorescent lamps belong to the family of low-pressure mercury gas discharge lamps. They are by far the most widespread discharge lamp types. They are available in both tubular and compact versions. Although the operating principle of compact fluorescent lamps is largely the same as that of the tubular version, their construction and performance are in many ways different from that of the tubular version [2, 3].

Working Principle

The discharge tube of a fluorescent lamp is filled with an inert gas and a little mercury and has an electrode sealed into each end (Fig. 1). To facilitate starting, the electrodes of most fluorescent lamps are preheated prior to ignition, which is accomplished by means of a high-voltage pulse, generated by an external device called the igniter or starter. (Igniters are available as simple glow-switch starters or as electronic devices.) When the lamp is switched on, the electrodes begin emitting electrons, and through the collision of these electrons with the gas atoms, the ionization process starts. The inert gas is then heated up, and the mercury inside the lamp is completely evaporated to give a mercury vapor pressure of about 0.8 Pa ($8 \cdot 10^{-6}$ atm). The emitted electrons collide with and excite the mercury atoms, resulting in the emission of ultraviolet radiation and a small amount of blue visible light. The inside of the discharge tube is coated with a mixture of fluorescent powders. The ultraviolet radiation is

Tubular and Compact Fluorescent Lamp,

Fig. 1 Main parts and principle of operation of a tubular fluorescent gas discharge lamp (TL) [1]



converted to visible light when it passes through the fluorescent powder coating.

Like almost all gas discharge lamps, a fluorescent lamp cannot be operated without some device to limit the current flowing through it. This device, usually in the form of an inductive coil, is called a ballast. Fluorescent lamps also need, again like most discharge lamps, an igniter for starting the lamp.

Tubular Fluorescent Lamps

Tubular fluorescent lamps are widely used in offices, schools, shops, and low-ceilinged industrial premises. So today their use is mainly indoors.

Materials and Construction

The main parts of a tubular fluorescent lamp are (see Fig. 1):

- Glass tube
- Fill gas
- Electrodes
- Fluorescent powder

Glass Tube

The tube of a normal fluorescent lamp is made of glass that is doped with a special material that blocks that UV radiation from the mercury discharge that is not converted by the fluorescent powder into visible light.

The original fluorescent lamp had a diameter of 38 mm. This tube diameter is usually characterized as T12, where the 12 stands for 12 times one-eighth of an inch (Table 1). This type of fluorescent tube is seldom seen today. During the late 1980s, a smaller and more efficient version was introduced that has a diameter of 26 mm. Today an even thinner and more efficient version has become the standard: the TL5 with a diameter of 16 mm. Even thinner fluorescent lamps are produced, but these are not for use in general lighting.

Fill Gas

The gas filling in a fluorescent lamp consists of a mixture of mercury vapor and an inert gas. The inert gas has three functions:

- To facilitate ignition, especially at lower temperatures
- To control the speed of the free electrons
- To prolong the life of the electrodes by reducing evaporation of electrode material

The inert gas usually consists of a mixture of argon and neon, although sometimes krypton is used as well.

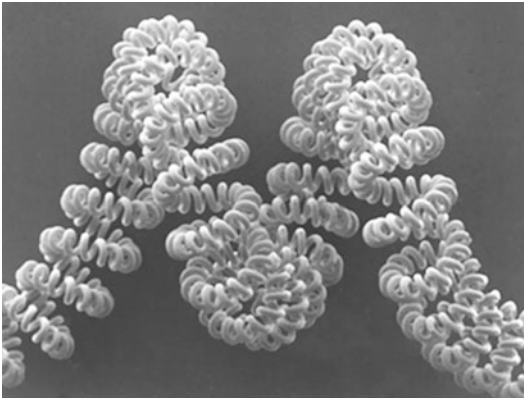
Electrodes

The function of the electrodes is to provide free-running electrons, which are necessary to start and maintain the discharge. A fluorescent lamp electrode consists basically of a tungsten filament (Fig. 2) that is coated with the so-called emitter material to facilitate electron emission.

Tubular and Compact Fluorescent Lamp,

Table 1 Fluorescent lamps with different tube diameters and their designations

Tube type	Diameter (inch)	Diameter (mm)
T12	12 * 1/8	38
T8	8 * 1/8	26
T5	5 * 1/8	16
T2	2 * 1/8	6
T1	1 * 1/8	2.8



Tubular and Compact Fluorescent Lamp, Fig. 2 Triple-coiled filament electrode [1]

Fluorescent Powder

The small crystals of the fluorescent powder applied to the inside of the discharge tube absorb the UV mercury radiation and convert it into visible light (this physical phenomenon is called luminescence). Different fluorescent powders convert the ultraviolet radiation into visible light of different wavelengths and thus different colors (Fig. 3).

By mixing different fluorescent powders in different proportions, lamps producing different tints of white light can be made. The type and composition of the fluorescent powder is the most important factor determining the light characteristics of a fluorescent lamp, such as color temperature, color-rendering index (R_a), and, to a large extent, the luminous efficacy of the lamp (lm/W). Some fluorescent powders convert the ultraviolet radiation into wavelengths covering almost the whole visible spectrum. Such powders, therefore, produce white light when used alone. However, their color rendering and efficacy are poor ($R_a < 70$ and lm/W < 80). Because these powders

lose some of their conversion activity relatively quickly, lamp-lumen depreciation is relatively large. Nowadays these lamps are hardly ever produced. Today fluorescent lamps often employ a mixture of three fluorescent powders, each having a very narrow spectrum band in red, green, and blue. In this way, white light is again obtained but of better color rendering and efficacy ($R_a > 80$, lm/W up to 105). The lamp-lumen depreciation with these powders is very low ($< 10\%$). For cases where extremely good color rendering is required, a mixture of more than three powders is used, resulting in lamps with excellent color rendering ($R_a > 90$), very low lumen depreciation, and high luminous efficacy (slightly smaller than the previous version, viz., up to some 90 lm/W).

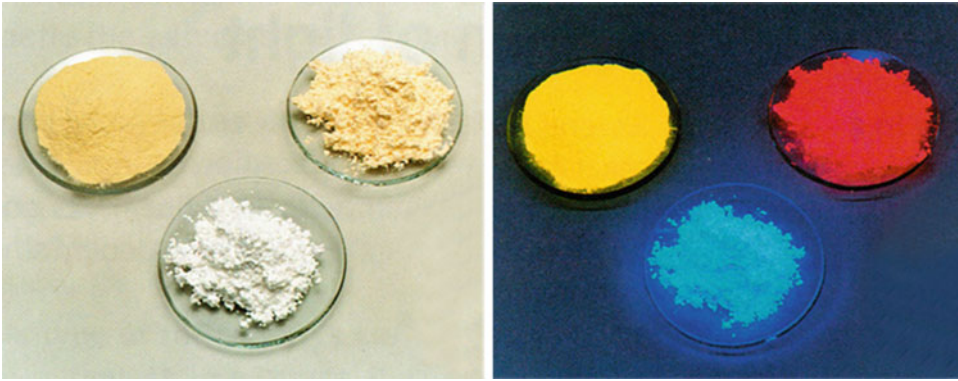
Properties

Energy Balance

Just below 30% of the input power is converted into visible radiation and a very small part into UV radiation. The rest is lost in the form of heat (at the electrodes, in the discharge itself, and as infrared radiation). Compare the figure of 30% with 8 and 12% for incandescent and halogen lamps, respectively.

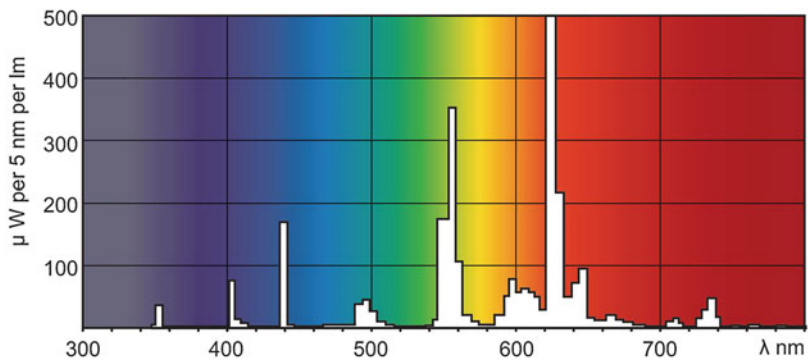
System Luminous Efficacy

As with most lamps, the luminous efficacy of tubular fluorescent lamps is dependent on the wattage of the lamp and the color quality of the light it gives. Types with color rendering better than R_a 80 and with low lumen packages have a system luminous efficacy starting at 50 lm/W, while lamps with higher lumen packages reach efficacies up to some 105 lm/W. Lamps with



Tubular and Compact Fluorescent Lamp, Fig. 3 Different types of fluorescent powder. *Left*, under white light, and *right*, the same fluorescent powders under UV radiation [1]

Tubular and Compact Fluorescent Lamp, Fig. 4 Spectral energy distribution of a fluorescent lamp color type 827: T_k 2700 K and R_a 80 [1]



extremely good color rendering ($R_a > 90$) have 15% lower efficacies.

Lumen-Package Range

Common tubular fluorescent lamps are produced in the range from some 500 to 6000 lm (corresponding wattage range approximately 8–80 W). Special very-high-output lamps are also produced in versions up to some 9000 lm (120 W).

Color Characteristics

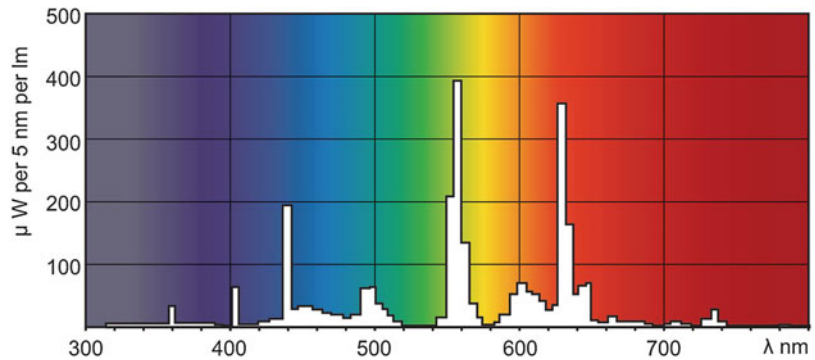
As has been mentioned, by mixing different fluorescent powders in different proportions, lamps with different spectra can be produced. As with all gas discharge lamps, the spectrum is discontinuous. Figures 4, 5, and 6 show the spectra of fluorescent lamps with different color temperatures T_k and color-rendering index R_a .

The color type designation used for fluorescent lamps is standardized, with the first digit standing for the color-rendering index R_a and the last two digits for the color temperature T_k . Thus, the color type 840 stands for a color-rendering index value R_a in the 80 ties and a color temperature of around 4000 K.

Present-day quality fluorescent lamps are produced in a whole range of color temperatures varying from 2700 K (warm white or incandescent lamp color tint) to 6000 K (bluish white), most of them in two color-rendering qualities with R_a in the 80 ties (800 series) and 90 ties (900 series), respectively. For special applications, versions are produced with extremely high color temperatures (up to some 17,000 K) with color rendering in the 80 ties. Today, fewer and fewer of those fluorescent lamps with poor color-rendering qualities (viz., R_a around 65 or less) that were produced in the past are produced.

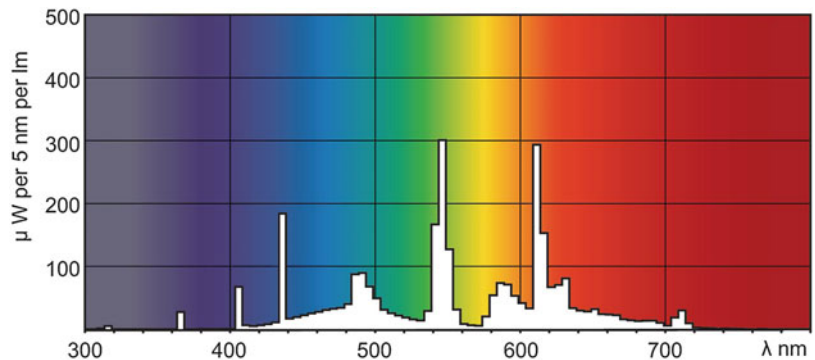
Tubular and Compact Fluorescent Lamp,

Fig. 5 Spectral energy distribution of a fluorescent lamp color type 840: T_k 4000 K and R_a 80 [1]



Tubular and Compact Fluorescent Lamp,

Fig. 6 Spectral energy distribution of a fluorescent lamp color type 940: T_k 4000 K and R_a 90 [1]



Lamp Life

During the operation of a lamp, the electrodes lose emitter material due to evaporation and as a result of the bombardment with ions from the discharge. This is the main cause of lamp failure: the lamp will no longer start, due to a broken electrode, or else it flickers because insufficient emitter material remains. The lifetime of fluorescent lamps and gas discharge lamps, in general, is thus very much dependent on the construction and materials of the electrodes. Conditions that influence lifetime are principally the type of gear used, the switching frequency, and the ambient temperature. The switching frequency plays a role because the high-voltage peak needed for ignition causes the electrodes to lose some of their material (sputtering). The more accurately the igniter-ballast combination preheats the electrodes, the less severe this effect is. The occurrence of such things as shocks or vibrations, different burning positions, and supply-power variations can play an additional role.

Today's high-quality fluorescent tubes have economic lifetimes (based on 20% mortality) of 15,000 h to more than 20,000 h if used on a so-called high-frequency or HF ballast and around 12,000 h when used on an electromagnetic ballast. For situations where the actions needed to replace lamps are expensive, special, and more expensive, long-life fluorescent lamp versions are available with lifetimes ranging from some 40,000 to 65,000 h. This long life is obtained by means of a special electrode design and construction.

Lamp-Lumen Depreciation

The main cause of lamp-lumen depreciation in a fluorescent lamp is that the fluorescent powder slowly becomes less active as a result of chemical attack by mercury ions. There may also be some blackening of the tube wall from the electrodes. For cool-white and cool-daylight lamp types, the lumen depreciation is higher than for lamps with a warmer tint. This is because of the faster depreciation of the blue fluorescent powders. Lamp-lumen depreciation for present-day high-quality

fluorescent tubes is around 10% after some 20,000 h.

Run-Up and Reignition

Fluorescent lamps operated on present-day electronic control gear start very quickly (within 1.5 s) and without flickering. This last is in contrast to lamps used on the older, glow-switch starters. After ignition, the light output increases quickly up to its maximum in about 1 min. When a fluorescent lamp is switched off, the vapor pressure drops so quickly that reignition is instantaneous.

Switching

It has already been mentioned that the switching frequency has an influence on lifetime, the influence being dependent on the type of electrical control gear used. Lamps operated with HF electronic-preheat-start ballasts show in general, little sensitivity to the switching cycle. This is because of the well-controlled starting conditions of the lamp (warm start). HF non-preheat starting leads to a relatively short lamp life under frequent-switching conditions.

Dimming

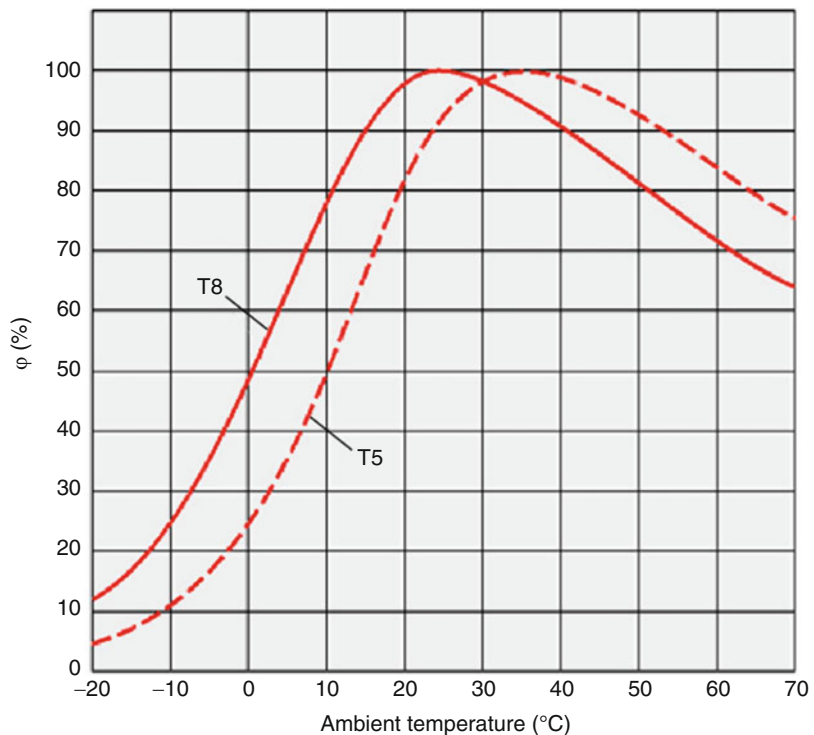
Dimming of modern fluorescent lamps on high-frequency electronic ballasts down to 3% of the nominal light-output value is easy, possible with an additional circuit that changes the operating frequency.

Ambient-Temperature Sensitivity

The luminous flux of a fluorescent lamp is determined by the mercury vapor pressure during operation. The mercury vapor pressure is in turn determined by the coldest spot in the tube, normally the tube wall, which of course is also dependent on the ambient temperature. The consequence of this is that the light output of fluorescent lamps is dependent on ambient temperature. The lumen output of fluorescent lamps is generally published for an ambient temperature of 25 °C. T12 and T8 lamps have their maximum light output at 25 °C, while the smaller diameter T5 lamps give their maximum light output at 35 °C (Fig. 7). To obtain a high- and near-constant light output over a wide temperature range, the mercury in some fluorescent lamp types is

Tubular and Compact Fluorescent Lamp,

Fig. 7 Relative light output of a T8 and a T5 fluorescent lamp in relation to the lamp's ambient temperature [1]



introduced into the tube not as a pure metal but as an amalgam. Such lamps have a lumen output that is near constant over the temperature range of 25–65 °C.

Mains-Voltage Variations

If the mains voltage varies, the power consumed by the lamp changes and with it the vapor pressure and consequently the lumen output. Given an optimized lamp/electromagnetic ballast combination, a 10% deviation in power will keep the lumen-output decrease usually at less than 10%. On high-quality, high-frequency electronic ballasts, a 10% deviation in power can keep the lumen-output decrease at less than 3%. In all cases, with higher mains-voltage decreases, the lumen output decreases rapidly.

Product Range

Tubular fluorescent lamps are being produced in a wide variety of types with different properties. The more important lamp types are listed in Table 2.

Up until 1990, fluorescent tubes were operated on electromagnetic ballasts. Even today, this way of operating a fluorescent tube is called “conventional,” although operation on high-frequency or HF electronic ballasts has since become the standard. T5 lamps can only be operated on HF electronic gear.

T5 lamps are produced in an execution where the design is such that the maximum luminous efficacy is obtained: the high efficiency, HE, and series of lamps. Another type of execution is optimized for maximum lumen output at the cost of efficacy: the high output, HO, and series (10% higher lumen output and 10% lower efficacy). Very-high-output lamps (VHO) are produced as well.

Some tubular fluorescent lamps are given a coating on the outside of the tube. A silicone water-repellent coating is applied to prevent starting problems for types that are used under conditions of high humidity. Types intended for use in food stores can be given a special protective coating that prevents surrounding products from

Tubular and Compact Fluorescent Lamp, Table 2 Different tubular fluorescent lamp types

Lamp aspect	Range
Tube diameter	38 mm (T12)/26 mm (T8)/16 mm (T5)
Lamp circuit	Electromagnetic/HF electronic
Balance efficacy/lm output	High efficacy (HE)/high output (HO)/very-high output (VO)
Color temperature T _k	Standard range (2700–6000 K)/8000–17,000 K
Color rendering R _a	R _a > 90 (900 series)/80 < R _a < 90 (800 series)/R _a < 80
Color	Red, green, blue, etc.
Lifetime	Standard (up to 20,000 h)/very long (up to 65,000 h)
Ignition method	Standard/low-temperature ignition
Tube coating	No coating/coating (silicon, protective, or reflector)
Safety of operation	Standard/explosion safe
Tube length	Standard (450–1800 mm)/mini (150–500 mm)
Tube shape	Tubular/U-shaped/circular

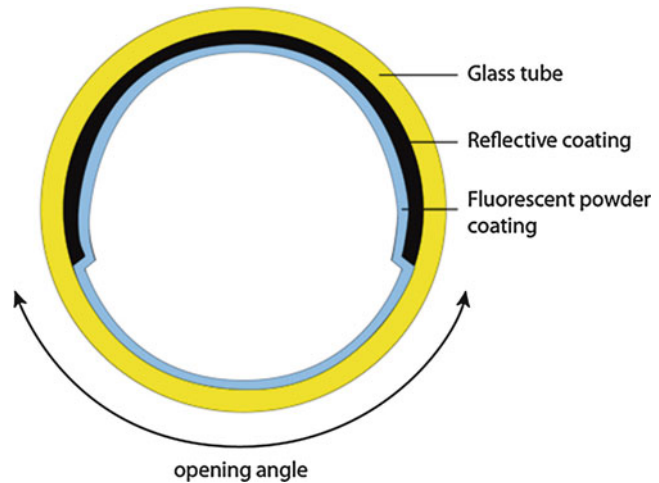
contamination in the event of accidental lamp breakage. Some of these latter types have a spectrum that makes the appearance of food more appealing (especially meat). In reflector fluorescent lamps, a diffusely reflective coating is applied between the upper part of the inner tube wall and the fluorescent powder. The light is concentrated through the uncoated area (or “window”) of the lamp so increasing the downward component of the light (Fig. 8).

Compact Fluorescent Lamps

Compact fluorescent lamps (CFLs) were initially developed (beginning of the 1980s) for use in those applications where incandescent lamps were traditionally used. Today, the application of compact fluorescent lamps has been widened and includes not only domestic lighting but office and road lighting (residential streets) as well. For the more compact versions, alternatives in the form of solid-state lamps are becoming increasingly more available.

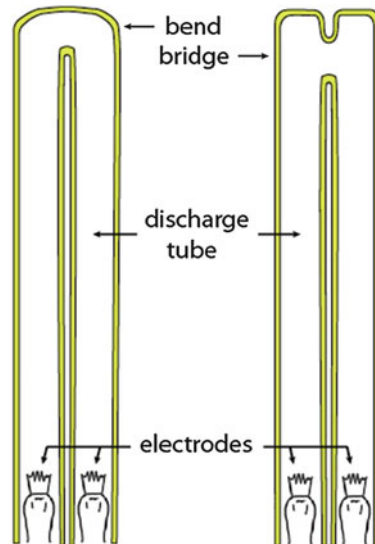
Tubular and Compact Fluorescent Lamp,

Fig. 8 Reflector fluorescent lamp



Working Principle

The gas discharge principle employed in compact fluorescent lamps is exactly the same as that in tubular gas discharge lamps (see previous section). Their compactness is achieved by reducing their length. This is done either by folding a longer tube into a shorter one or by joining together two or more parallel tubes so that one open pathway is obtained where free electrons and ions can move from one electrode to the other, as in a normal straight fluorescent tube (Fig. 9). The folding can be repeated, or the interconnection can be done with more than two parallel tubes (always keeping one open pathway) to further increase the size of the lamp.



Materials and Construction

As far as fill gas, electrodes, and fluorescent powders are concerned, compact fluorescent lamps are essentially the same as conventional tubular fluorescent lamps. Reference is therefore made to the previous section “[Tubular Fluorescent Lamps](#).”

The main difference in construction between compact and tubular fluorescent lamps lies in the tube shape and the lamp caps. But another point of difference is that in those compact lamps that need to work directly from the mains without any external electrical component, the igniter and ballast have to be integrated in the lamp itself. In this

Tubular and Compact Fluorescent Lamp, Fig. 9 Folding one tube (*left*) or connecting two separate tubes (*right*) to form a compact fluorescent lamp with one open pathway between the electrodes

case, the lamp foot is used for housing the control gear.

Tube

A great variety of tube shapes are produced. Figure 10 shows versions where tubes are interconnected. All these are examples of compact fluorescent lamps where the gear is not integrated in the lamp itself (compact fluorescent lamps non-integrated CFL-NI). Figure 11 shows a set of



Tubular and Compact Fluorescent Lamp, Fig. 10 Nonintegrated two-leg compact fluorescent lamps. From *left to right*: two pins with bridge connection, two pins with bend connection, and four pins with bridge connection



Tubular and Compact Fluorescent Lamp, Fig. 11 Triple-folded nonintegrated compact fluorescent lamps



Tubular and Compact Fluorescent Lamp, Fig. 12 Single-plane-folded nonintegrated compact fluorescent lamps

triple-folded types, while Fig. 12 shows two different folded, single-plane versions.

Figure 13 shows examples of compact lamps where the gear (HF electronic) is integrated in the lamp foot (integrated CFL-I). Since these lamps can directly replace incandescent lamps, it is also desirable that their shape and dimensions should be very close to those of normal incandescent lamps. Figure 14 shows two retrofit versions where an outer bulb with an internal diffusing coating also ensures that the light distribution is close to that of an incandescent lamp.

Lamp Cap

The compact fluorescent lamps with integrated control gear have the same lamp caps as normal incandescent lamps, viz., Edison screw type and bayonet caps.

In most nonintegrated, small, twin-tube versions, the starter is built into the lamp cap itself (integrated starter but nonintegrated ballast). All nonintegrated types are fitted with special caps. A large variety of caps and bases exist in order to ensure that only the correct type of lamp can be used in a given situation (especially defined by the type of gear used in the particular luminaire).

The lamp caps of nonintegrated compact fluorescent lamps are of the push-fit type (they fit by pushing them into the lamp holder). They are available in both square-shaped and rectangular-shaped versions. Lamp caps for lamps with an integrated starter employ two-pin connectors, while lamps for use with electronic control gear



Tubular and Compact Fluorescent Lamp, Fig. 13 Differently shaped integrated compact fluorescent lamps



Tubular and Compact Fluorescent Lamp, Fig. 14 Retrofit compact fluorescent lamps with shape, dimensions, and light distribution close to that of an incandescent lamp bulb

or dimmers have four-pin connectors (Fig. 10 right, Fig. 11 right, Fig. 12 right).

Properties

Energy Balance

Approximately 20% of the input power of compact fluorescent lamps is emitted in the form of visible radiation. A tubular fluorescent lamp emits

some 28% of visible radiation. The difference is mainly due to the fact that in compact lamps, the multitude of closely packed tube parts absorbs part of the light.

System Luminous Efficacy

The luminous efficacy of a compact fluorescent lamp is partly dependent on the wattage of the lamp and the color quality of the light emitted but more so on how and how many times the compact tube is folded. In low-wattage versions, good color-rendering versions have a system luminous efficacy starting at 45 lm/W, while higher-wattage versions reach efficacies of up to some 70 lm/W.

Lumen-Package Range

Compact fluorescent lamp with integrated control gear is produced in the range from some 250 to 2000 lm (corresponding wattage range approx. 5–35 W). Nonintegrated types are available in the range from some 250 to 6000 lm (corresponding wattage range approx. 5–80 W).

Color Characteristics

The color characteristics of compact fluorescent lamps are principally the same as those of normal fluorescent lamps. Reference on this point is therefore made to the previous section “[Tubular Fluorescent Lamps](#).” The range of different color temperatures and color-rendering indices available with compact fluorescents is in practice

somewhat limited compared to that found with the very wide range of tubular lamps.

Lamp Life

The lifetime of compact fluorescent lamps is very long compared with that of incandescent lamps, although it is usually shorter than that of tubular fluorescent lamps. Integrated versions (CFL-I) have, depending on the type, a rated average lifetime of between 8000 and more than 15,000 h. Note that the life of retrofit integrated CFLs is usually specified as rated average life (50% mortality), as is the case with incandescent lamps (which have an average life of 1000 h). Again, depending on the type, nonintegrated versions (CFL-NI) have economic lifetimes (based on 20% mortality) of 7000–20,000 h.

Lamp-Lumen Depreciation

The lamp-lumen depreciation of compact fluorescent lamps is similar to that of tubular fluorescent lamps. See previous section “[Tubular Fluorescent Lamps](#).”

Run-Up and Reignition

There are no fundamental differences here compared with normal fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Switching

There are no fundamental differences here compared with normal fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Dimming

Most integrated compact lamps are not dimmable. However, special, more expensive versions, with standard Edison or bayonet cap, are produced that can be dimmed to approximately 5% of full light output. Dimming of nonintegrated, four-pin, compact lamps is possible and is fundamentally the same as with tubular fluorescent lamps.

Ambient-Temperature Sensitivity

With each different lamp shape, the location of the coldest spot is also different, and this has consequences for the mercury pressure and therefore for the ambient-temperature sensitivity. With some shapes, the coldest spot is also influenced by the

Tubular and Compact Fluorescent Lamp, Table 3 Lamp properties versus lamp types

Lamp property	Range of lamp types
Gear integration	Integrated (I)/nonintegrated (NI)
Tube diameter	13 mm/18 mm
Shape	Various
Lamp length	80–200 mm (CFL-I)/100–600 mm (CFL-NI)
Lamp circuit	Electromagnetic/HF electronic
Color temperature T_k	Standard range (2700–6000 K)
Color rendering R_a	$R_a > 90$ (900 series)/ $80 < R_a < 90$ (900 series)

burning position of the lamp (e.g., base up or base down). The folded, integrated versions often use amalgam instead of pure mercury to stabilize the mercury pressure and thus to minimize the temperature sensitivity.

Mains-Voltage Variations

No fundamental differences compared with tubular fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Product Range

Compact fluorescent lamps are being produced in many types with different properties. The more important lamp properties for which different versions are made are gear integration (integrated or nonintegrated), tube diameter (13 mm and 18 mm), shape, tube length (related to lumen package), lamp circuit, color temperature, and color rendering. Table 3 indicates for all these lamp aspects the range of lamp types available.

Cross-References

► [Phosphors and Fluorescent Powders](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)

2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. IESNA, New York (2011)

Tungsten Filament Lamp

- [Incandescent Lamp](#)

Turkey Red

- [Alizarin](#)

Typical Color

- [Memory Color](#)

U

UCS Diagram

- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

Ulbricht Sphere

- [Integrating Sphere](#)

Ultrafast Imaging

- [Transient Imaging](#)

Ultraviolet Radiator

- [Blacklight Lamp](#)

Umov Effect

Ken Tapping
D.R.A.O, National Research Council, Penticton,
BC, Canada

Synonyms

[Umov's law](#)

Definition

When light is scattered off particles in a dusty surface or cloud, it becomes polarized. The more reflective (higher albedo) the dust particles, the lower the overall degree of polarization in the scattered light. This inverse relationship was first noted by Nikolay Umov and has accordingly become known as the “Umov effect.”

Introduction

When randomly polarized light is obliquely incident on particles in a dusty surface or layer, the scattered or reflected light becomes partially polarized to a degree that is dependent upon the viewing direction with respect to the incident light (phase angle) and, if the particles are irregularly shaped, their orientation. Since the geometry of the source-scatterer-observer light path is the same for all single-particle scattering events, this component of the light is polarized. For light scattering involving a large number of particles, the geometries are many and widely different, so that even though each scattering event might produce a net polarization, the integrated scattered light from these particles will be randomly polarized. The higher the reflectivity of the individual particles, the larger the contribution to the randomly polarized light reaching the observer from the multiple-particle scattering events. This means the higher the general reflectivity (albedo) of the

dust particles, the lower the net degree of polarization of the light reaching the observer. This relationship is described by the following inverse correlation:

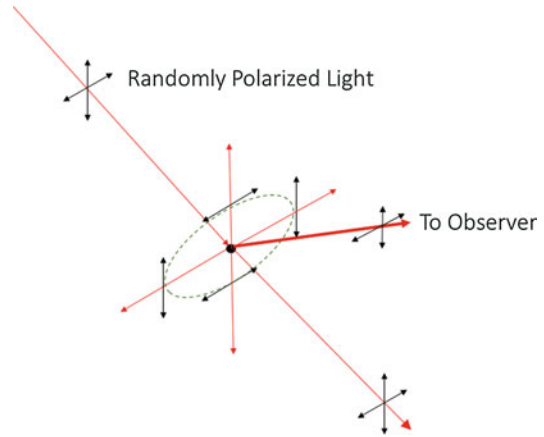
$$P \propto \frac{1}{A}$$

where P is the degree of polarization and A the overall (geometric) albedo of the object. Consequently, light scattered by dim, dusty objects will be more polarized than light from bright, dusty objects.

The relationship was first described by Russian physicist Nikolay Umov in 1905 [2] and has become widely referred to as the **Umov effect** and the inverse relationship between maximum degree of polarization and geometric albedo as **Umov's Law**. It is usually described as giving a straight line with negative slope in a log-log plot of $\log(P_{\max})$ versus $\log(A)$, in a so-called linear **Umov diagram**. The degree and directional properties of the scattered light depend upon the shapes and sizes of the scattering particles and are a useful tool for investigating dusty surfaces such as the surface of the Moon or asteroids or even cosmic and interplanetary dust clouds.

One Light Ray, Scattering Off a Single Particle

Figure 1 shows a beam of randomly polarized light coming from the far background in the upper left propagating toward the lower right foreground. It is incident upon a particle, which is assumed to be “small” and “spherical,” and is scattered. The light that continues with its direction unchanged retains its initial polarization. However, in general the scattered light changes its polarization to a degree that depends on the direction in which it is scattered. Light deviated by 90 degrees with respect to the direction of the incident light becomes totally linearly polarized in a sense tangential to a circle centered on the particle and normal to the direction of the incident ray. In an arbitrary direction (the thick line in the figure), the light will be partially polarized to



Umov Effect, Fig. 1 Polarization of light by scattering off a single particle

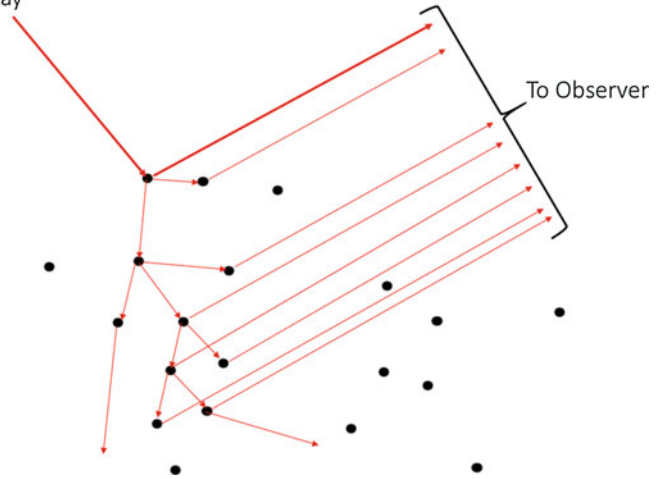
some degree depending upon the scattering angle. If the illumination source and the observer are a large distance from the scatterers, light from a collection of particles reaching the observer after a single interaction would follow the same path geometry, and their individual polarization components would be additive and give a calculable increase in the degree of polarization seen by the observer.

In the diagram the black arrows indicate relative degrees of polarization in the horizontal and vertical directions. Totally linearly polarized light will have just one component. Unpolarized light will have its orthogonal components of equal magnitude. In general there is some degree of polarization, in which case the longer arrow indicates the dominant plane of polarization.

Scattering in an Ensemble of Many Particles

In general, when incident light encounters a dust cloud or dusty layer, it interacts with a very large number of particles. In this discussion, the term “particle” refers to an “average particle,” combining the scattering properties of the ensemble of particles. Each of these scatters light onto an even larger number of particles, which, in turn, give rise to further scattering. This makes it possible to consider an average scattering event, involving

Umov Effect, Fig. 2 A ray of light interacting with a single particle and triggering many cascades of multiple particle scattering



an average cascade of secondary scattering, leading to the following simplifying approximations:

1. The total scattering process in the cloud or dust layer can be treated as a very large number of average scattering cascades, added together.
2. A very large number of irregular, randomly oriented particles can be treated, on average, as spherical, isotropically scattering particles in a scattering cascade. Each particle is assumed to have an individual albedo a , which is the same magnitude in all directions.
3. For scattering paths involving only one particle between the light source and the observer, the geometries will be identical and all contributing the same polarization components to the observer, so these polarizations will be additive.
4. The geometries of the secondary scattering events will differ widely and can be assumed random. We assume here there is a very large number of these that will all contribute different polarization effects to the light sent in the direction of the observer, so collectively these will average out and contribute to a randomization of the polarization.
5. That every ray of light ultimately encounters a particle unless it is directly scattered from the source in the direction of the observer.

This scattering scenario is represented in Fig. 2.

Albedo

Light from the source hits a particle. A fraction a of this is scattered in the direction of the observer. This fraction depends on two main factors, firstly the reflectivity of the particle and secondly the geometry of the interaction, i.e., that the light is now being scattered in all directions. We assume that apart from the light directly scattered in the direction of the observer, all rays of light eventually encounter a particle. Then if each scattering event illuminates n particles on average, and each particle sends some light in the direction of the observer, the observed geometric albedo, A , of the cloud or dust layer is given approximately by:

$$A \propto a + a^2 n + a^3 n^2 \dots a^k n^{k-1} \dots \quad (1)$$

where $an < 1$ and we assume the scattering process proceeds ad infinitum, although the contributions to the redirected light eventually become vanishingly small, so the equation becomes:

$$A \propto \frac{a}{1 - an} \quad (2)$$

This geometric albedo, A , is a relative property that refers to the ratio of the reflectance of the

particle to that of a perfectly reflecting Lambertian surface with the same area.

Polarization

The first encounter between the light and a particle has the same geometry as all the other first encounters. However, the large number of ray paths of the encounters with second and subsequent particles will contribute a net randomization of the polarization. We define the degree of linear polarization, P , by:

$$P = \left| \frac{I_y - I_x}{I_y + I_x} \right| \quad (3)$$

where x and y are orthogonal directions with respect to the direction of the scattering plane defined by the location of the source-scatterer-observer. In this case P varies between zero for a randomly polarized (unpolarized) wave and unity for one that is fully linearly polarized.

Since the observer sees the sum of a very large number of cascades, we can assume the first scatter in each cascade yields some level of linear polarization P^* say, but all subsequent scatterings, with their various scattering geometries, yield randomly polarized radiation, the net observed polarization will be weighted for the albedo-related decreases as the cascade is descended:

$$P \propto \frac{P^* + 0 + 0 + 0 + 0 \dots}{a + a^2n + a^3n^2 + \dots a^kn^{k-1} \dots} = P^* \left(\frac{1 - an}{a} \right) \quad (4)$$

So substituting from the albedo eq. (2) where P^* is a constant for the particle population, we have that the observed polarization is decreased in proportion to the contribution of increased interparticle scattering that increases albedo:

$$P \propto \frac{1}{A} \quad (5)$$

This is “Umov’s Law,” and the inverse relationship, where the linear polarization of the scattered light is a maximum for weakly

absorbing, low albedo particles, is known as the “Umov effect.”

Practical Uses of the Umov Effect

The scenario described here involves the integrated contributions of a large number of quasi-random or at least unpredictable events, and according to the central limit theorem, fluctuations in the amplitude and polarization have the properties of a Gaussian process. If the scatterers are non-spherical, but their orientations are random, and there are many of them, the results of the calculation would be similar. Under these conditions it is possible to detect in some cases differences in surfaces, such as in lunar soils, but little can be inferred about the scattering media involved.

However, when the particles are fewer in number or the cascade of scattering is not extensive (e.g., the scatterers have a low individual albedo), the probability density functions of individual fluctuations will deviate from a Gaussian distribution, although they might integrate to this normal distribution, given sufficient time. Some inferences can be made about the scattering particles from observations, but in many cases, the easiest approach is through laboratory measurements and simulations aimed at duplicating the observations. The practical value of the Umov effect has been primarily for evaluating the size of distant celestial objects that appear as a point source to the observer [3].

The Umov effect has also proved useful in investigating the scattering particles producing the zodiacal light, i.e., interplanetary dust particles [5], and the regolith, i.e. dust, gravel, and small rock fragments covering solid rock, on asteroids and the Moon. Other studies include terrestrial aerosols and pharmaceuticals, paints and coatings, and remote sensing through scattering media. It has also been shown to provide a quantitative estimation of the reflectance of small irregularly shaped particles in a two-component mixture, which are commonly found in stratospheric aerosols, desert sands, and thin clouds of cosmic dust [4].

Cross-References

- [Apparent Magnitude, Astronomy](#)
- [Instrument: Photometer](#)
- [Interactions of Color and Light](#)
- [Light, Electromagnetic Spectrum](#)
- [Rayleigh and Mie Scattering](#)

References

1. Merzbacher, E., Feagin, J.M., Wu, T.-H.: Superposition of the radiation from n independent sources and the probability of random flights. *Am. J. Phys.* **45**, 964 (1977)
2. Umov, N.: Chromatische Depolarisation durch Lichtzerstreuung. *Phys. Z.* **6**, 674–676 (1905)
3. Zubko, E., Videen, G., Shkuratov, Y., Muinonen, K., Yamamoto, T.: The Umov effect for single irregularly shaped particles with sizes comparable with wavelength. *Icarus*. **212**, 403–415 (2011)
4. Zubko, E., Videen, G., Subko, N., Shkuratov, Y.: The Umov effect in application to an optically thin two-component cloud of cosmic dust. *Mon. Not. R. Astron. Soc.* **477**, 4866–4873 (2018)
5. Zverev, A.: The Umov Effect: space dust clouds and the mysteries of the universe – FEFU scientists are developing and methodology to calculate the ratio of dust and gas in the comas and tails of comets. *Eur. Secur.* (2018)

Umov's Law

- [Umov Effect](#)

Understanding Perception and Expression of Color in Ancient Languages

- [Evolution of Color Expression](#)

Uniform Chromaticity Scale Diagram

- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

Unique Hues

- [Primary Colors](#)

Universalism-Relativism Debate

- [Color Categorization: Patterns and Mechanisms of Evolution](#)

USC Diagrams; Uniform Chromaticity Scales; $Y_u'v'$

Stephen Westland
Colour Science and Technology, University of
Leeds, Leeds, UK

Definition

The coordinates of the CIE 1960 UCS diagram are calculated from the 1931 CIE XYZ tristimulus value or from the xy chromaticity coordinates thus:

$$\begin{aligned} u &= 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3), \\ v &= 6Y/(X + 15Y + 3Z) = 6y/(-2x + 12y + 3). \end{aligned} \quad (1)$$

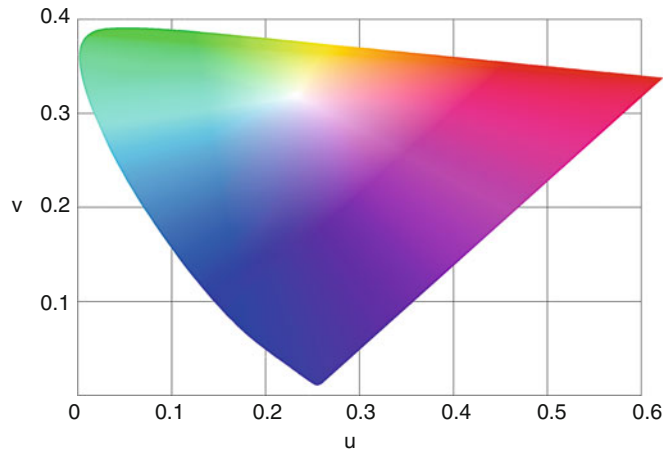
The 1976 CIE $u'v'$ chromaticity diagram was obtained by stretching the v -axis of the UCS diagram ($u' = u$, $v' = 1.5v$), and the coordinates are calculated thus:

$$\begin{aligned} u' &= 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3), \\ v' &= 9Y/(X + 15Y + 3Z) = 9y/(-2x + 12y + 3), \end{aligned} \quad (2)$$

where, alternatively, $U' = 4X/9$, $V' = Y$, and $W' = -X/3 + 2Y/3 + Z/3$ and $u' = U'/(U' + V' + W')$ and $v' = V'/(U' + V' + W')$.

USC Diagrams; Uniform Chromaticity Scales;

Y_u/v' , Fig. 1 CIE 1960 UCS diagram. Figure by Adoniscik (Own Work) [Public Domain] via Wikimedia Commons



Both the CIE 1960 UCS diagram and the 1976 CIE $u'v'$ chromaticity diagram are associated with the 1931 Y tristimulus value to provide a complete trichromatic specification since $V = V' = Y$.

Historical Development

The 1931 CIE system of colorimetry [1] allowed color stimuli to be defined in terms of tristimulus values XYZ but does not provide a particularly uniform representation of color stimuli in visual terms. If lines are drawn on the CIE 1931 chromaticity diagram that represent equal perceptual steps, then the disparity in the lengths of the lines is as great as 20 times, with the lines in the green region, for example, being much longer than those in the blue region [2].

However, many years earlier it had been suggested, by König, for example, that “it ought not to be difficult in the construction of a chromaticity diagram so to modify the adopted arbitrary assumptions that the separation of two points on it would give a measure for the difference in sensation between the colors corresponding to them” [3]. Following work by Judd, in 1935 a Maxwell triangle yielding approximately uniform chromaticity scales (known as the UCS diagram) was defined by reference to the CIE system, and in 1937 MacAdam developed a rectangular coordinate system based on the simple transformation of tristimulus values. In 1959 at a meeting in

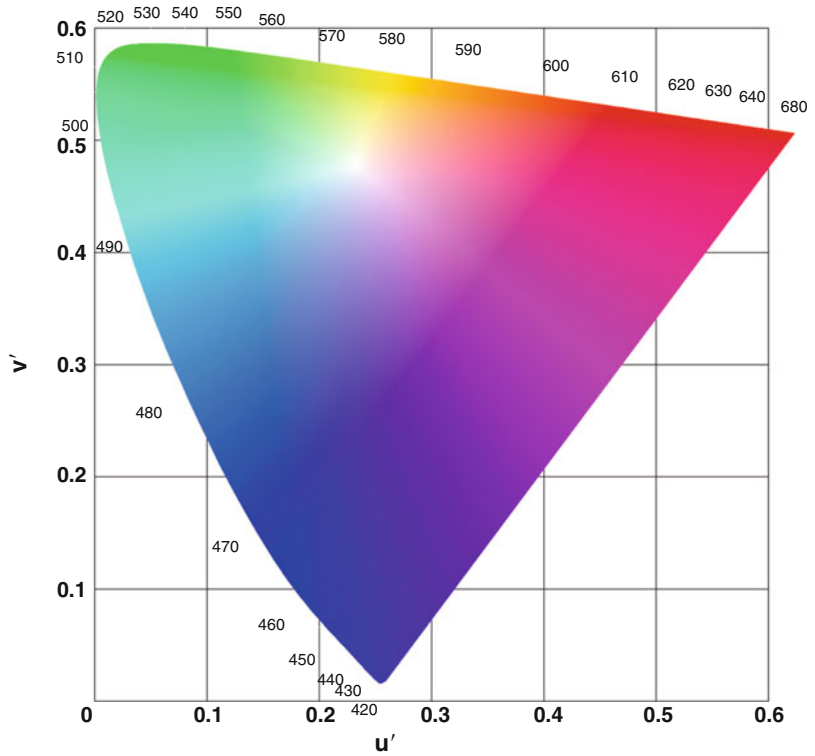
Brussels, the CIE recommended the 1937 MacAdam uv -diagram for use whenever a projective transformation of the CIE xy -diagram is desired to give uniform chromaticity spacing [4]. This space became known as the CIE 1960 UCS diagram (Fig. 1).

If compared with the 1931 chromaticity diagram, the effect of the linear transformation was to elongate the blue-red portions of the diagram and to relocate the white point to reduce the size of the green area. Although not perfectly uniform, the UCS diagram was thought to be an almost optimal transformation from the 1931 space [2] and was deemed to be sufficient for most practical purposes [4]. However, at a meeting in London in 1975, the CIE proposed modifying the UCS diagram and replaced it with the $u'v'$ -diagram stretching the v -axis ($v' = 1.5v$). The resulting diagram was adopted as the CIE 1976 UCS diagram (Fig. 2).

Properties and Current Status

All chromaticity diagrams, whether xy , uv , or $u'v'$, have the property that additive mixtures of colors are represented by points lying on the straight line joining the points representing the constituent colors [5]. However, the CIE 1960 UCS diagram and the 1976 CIE $u'v'$ diagram represent substantial improvements over the 1931 CIE xy chromaticity diagram in terms of the visual uniformity of

USC Diagrams; Uniform Chromaticity Scales; $Y_u'v'_u'$, Fig. 2 CIE 1976 UCS diagram. (Figure by Adoniscik (Own Work) [Public Domain] via Wikimedia Commons)



the spaces. Whereas lines that represent equal perceptual steps drawn on the CIE 1931 chromaticity diagram differ in length by as much as 20 times, the same lines drawn on the CIE 1976 $u'v'$ diagram differ in length by about four times (and over much of the diagram, the difference is not greater than two to one) [2].

CIE 1976 UCS is useful for showing the relationships between colors whenever interest lies in their discriminability. However, it is mainly used for the representation of self-emissive colors on display devices or those produced directly from light sources since the diagram assumes that colors are of equal luminance (this condition is rarely met in practice with reflective surface colors). Note, however, that CIELUV, the first approximately uniform three-dimensional space, is a transformation of the CIE 1976 UCS chromaticity coordinates u' , v' , and Y .

Cross-References

- [CIE Tristimulus Values](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [CIELAB](#)

References

1. CIE Pub. No. 15: Colorimetry. Central Bureau of the CIE, Vienna (2004)
2. Hunt, R.W.G.: The Reproduction of Colour, 6th edn. Wiley, Chichester (2004)
3. Judd, D.B., Yonemura, G.T.: CIE 1960 UCS diagram and the muller theory of color vision. J. Res. Natl. Bur. Stand. A. Phys. Chem. **74A**(1), 23–30 (1970)
4. Wright, H.: The Measurement of Colour, 4th edn. Adam Hilger, London (1969)
5. Hunt, R.W.G., Pointer, M.R.: Measuring Colour, 4th edn. Wiley, Hoboken (2011)

V

$V(\lambda)$, $V'(\lambda)$

► Spectral Luminous Efficiency

Vagueness

► Color Categories and Vagueness

Vantage Theory of Color

Adam Głaz

Department of Modern Languages, Maria Curie-Skłodowska University (UMCS), Lublin, Poland

Synonyms

VT and color

Definition

Vantage Theory (henceforth VT) is a cognition-based model of color categorization proposed by a one-time student of Brent Berlin and Paul Kay, Robert E. MacLaury [1, 6, 8–11]. The model's major tenet is that humans construct color categories as one, two, or occasionally three vantages

(i.e., points of view), a category being an assembly of its vantages.

Vantage Theory Step by Step

Overview

Vantage Theory was proposed after MacLaury and his coworkers had conducted interviews with about 900 speakers of 116 Mesoamerican languages (Mesoamerican Color Survey, part of World Color Survey), later enriched with data from a wide spectrum of world languages. The interviews consisted of three procedures – *naming*, *focus selection*, and *mapping* – and were performed with the use of the Munsell set of color chips, 320 chromatic and 10 achromatic (Fig. 1).

First, in the procedure of naming, the informant was shown the chips one by one in random order and asked to name each. The naming ranges of each color term were then marked on the derandomized array. Next, the informant was asked to choose the focus (best example) of each color term used. Finally, he/she was shown the arranged set without the naming ranges and asked to indicate all chips he/she would refer to with a given term. This process of mapping proceeded in incremental steps, until the informant refused to continue. Thus, the naming range, the focus/foci, and the mapping range of each term were elicited.

Parallelisms between the categorizing behavior of informants and spatiotemporal orientation were



Vantage Theory of Color, Fig. 1 The Munsell color array. The horizontal axis hue, from red to purple, severed in the middle of red at column 40. The vertical axis is for

brightness. All chips at maximum saturation. On the left: ten achromatic colors, from white to black. (© Hale Color Consultants, Inc. Reproduced with permission)

Vantage Theory of Color, Fig. 2 Modeling of the blue category in VT. *Bu* blue focus, *S* attention to similarity, *D* attention to difference

Levels	Fixed Coordinates	Mobile Coordinates	Entailments
1	Bu	S	focus, range
		↗	
2	S	D	breadth, margin

observed, the analogy being drawn in an instinctive and neurally expedited manner. While constructing a color category, a person anchors their cognition in a given dimension of color (hue, brightness, or saturation) and relates color stimuli to that fixed coordinate through similarity or its lack. The fixed coordinate is a categorical equivalence of spatial landmarks, whereas attention to similarity or difference arises by analogy to experiencing relative motion. Constructing a color category is here illustrated with a hypothetical example in Fig. 2.

On level 1, the category is endowed with the blue focus, the starting point for category construction, and a range of color stimuli similar (*S*) to the category focus. When the categorizer starts emphasizing difference (*D*) more than similarity, the category is curtailed at a margin. Through an analogy to motion, VT defines color similarity and difference as reciprocal and gradable, with the endpoints of the cline being total identity and complete disparity. Speakers can shift their attention by moving between the two extremes. Hue (most commonly), brightness (less so), or saturation (rarely) can function as *inherently fixed*

coordinates, whereas *S* and *D* are *inherently mobile*: each type can become the other for immediate purposes of constructing a figure-ground arrangement of coordinates (where the fixed coordinate is the ground or “given” and the mobile coordinate is the figure or “new”). This, however, does not deprive either type of its fixed or mobile status, respectively. *S* and *D* are indispensable, whereas the inherently fixed coordinates depend on the domain of categorization. VT emphasizes hue, brightness, and saturation as important color dimensions based on the implied salience of these perceptual features that is typically found in perceptual experience. If the salience of these dimensions were nonuniform, or if other dimensions took precedent over hue, brightness, and saturation experience, then VT analyses could still be applied to these alternative constructs in a similar fashion.

The ground-to-figure arrangement, in which a mobile coordinate is fixated to serve as ground for the introduction of a new value, is called a *vantage*. Crucially for the categorizing process, there may be two or sometimes three vantages (points of view) on a category, the category being the

DOMINANT VANTAGE			RECESSIVE VANTAGE			
Entailments	x			y		Entailments
	Fixed Coordinates	Mobile Coordinates		Fixed Coordinates	Mobile Coordinates	
focus, range	Bu	S	1	Gn	D	focus, margin
		↗			↗	
breadth	S	Gn	2	D	Bu	curtailment
		↗			↗	
margin	Gn	D	3	Bu	S	range

Vantage Theory of Color, Fig. 3 Modeling of the cool category in VT

Vantage Theory of Color, Table 1 Characteristics of the dominant and recessive vantages. Elemental colors are “the purest, most intense perceptions” of red, yellow, green, blue, white, and black [8, p. 467], although black

and white are traditionally not treated as colors. Elemental colors have their specific locations in the Munsell array, cf. Fig. 4

Dominant vantage	Recessive vantage
Greater number of chips covered	Smaller number of chips covered
Range more concentrated – over a more compact area	Range more dispersed – over a larger area
Focus more centralized relative to elemental colors	Focus less centralized relative to elemental colors

assembly of the vantages that compose it. In space-time, the same event involving motion is perceived differently depending on one’s location (e.g., on a train vs. while standing by the track). As an example of point of view in color, consider the hypothetical cool category in Fig. 3, named with two terms, *x* and *y*. The two points of view are the blue-focused similarity-based *dominant* vantage and the green-focused difference-based *recessive* vantage. They are characterized by a reversal of coordinates that swaps an emphasis on similarity to the focus with an emphasis on difference from the focus.

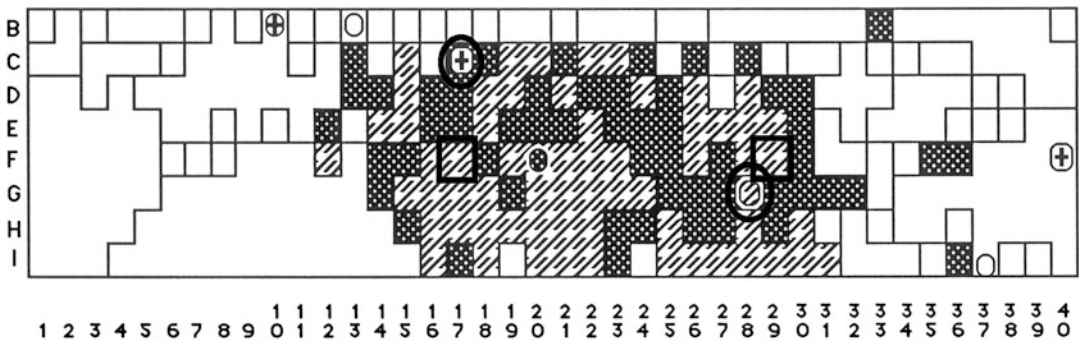
In the blue-focused dominant vantage, the emphasis on *S* is juxtaposed with *Gn* (green), the second focus of the category, which is then juxtaposed with *D*. The recessive vantage starts with the *Gn* focus. However, here the categorizer first establishes the boundary of the vantage by concentrating on difference vis-à-vis a blue category focus, *Bu*, which is introduced as a mobile coordinate on level 2 but returns to its inherently fixed capacity on level 3. There, it is treated as (weakly) similar to *Gn*, the process defining the range of the vantage.

Together, two vantages as points of view are the way VT models the complexity of categories that is seen in the empirical data. According to VT analyses, in such data one can explain the existence of cool or warm categories as the cognitive trading off of emphases of similarity and difference.

The most important differences between the two vantage types are listed in Table 1.

Figure 4 illustrates these differences with the cool category in Zulu.

The dominant *hlaza* names 62 chips, as opposed 56 for the recessive *kosazana*, but it only spans 20 columns vs. 25 that *kosazana* does. *Hlaza* is focused in G28, which is very near elemental blue in F29, whereas *kosazana*’s first focus (the second focus being disregarded for the present purposes) falls on C17, as many as three rows above elemental green in F17 (interestingly, elemental green in F17 is named with the blue-focused *kosazana*). These visible effects (entailments) of the two points of view on the category are like a person’s spatiotemporal location being determined through reference to stationary and moving objects.



Vantage Theory of Color, Fig. 4 Naming and focusing of the cool category in Zulu (Bantu group, Niger-Congo family). *Lighter hatching*, blue-focused dominant *hlaza*; *darker pattern*, green-focused recessive *kosazana*. *Dark*

squares: elemental blue (F29) and elemental green (F17). *Dark ovals*: foci for *hlaza* (G28) and *kosazana* (C17). (Figure received by the author from Robert E. MacLaury)

Semantic Relations Between Vantages

In addition to modeling color dimension relations as just described, there are also three major types of semantic relations between the dominant and recessive vantages (*near synonymy*, *coextension*, and *inclusion*) that influence the ways color is ultimately categorized. These are “segments of a continuum, not discrete kinds of relation” [8, p. 112].

In near synonymy, the vantages are very much alike in terms of focus selection and range, and the differences are minimized. An example is the warm category, termed *he* and *lu*, constructed by a speaker of Jicaque (or Tol), an isolate spoken in Honduras [8, p. 123].

Coextension is a unique kind of relation and requires more attention. The fullest account can be found in [8]; see also [6, 9, 10]. Coextension was first observed in the warm category of Uspantec (Uspanteco), a Mayan language of Guatemala, and later in many interviews in Mesoamerica and elsewhere. The first four characteristics below are more common than the remaining ones, though all are subject to some degree of variation:

1. One category is named with two different root terms.
2. Each of the two terms is focused in reference to a different elemental hue.
3. The mapping of each term encompasses the focus of the other.

4. There is substantial overlap of the mapping of the two terms.

The more variable features are:

1. Mappings of the two ranges progress in opposite directions.
2. Naming ranges are intermixed, so that chips named with one term may be surrounded by those named with the other.
3. Foci of one or sometimes both terms are polarized, moderately when a term is focused between the category margin and the relevant elemental hue, in extreme cases when the focus falls outside the naming range of the term.

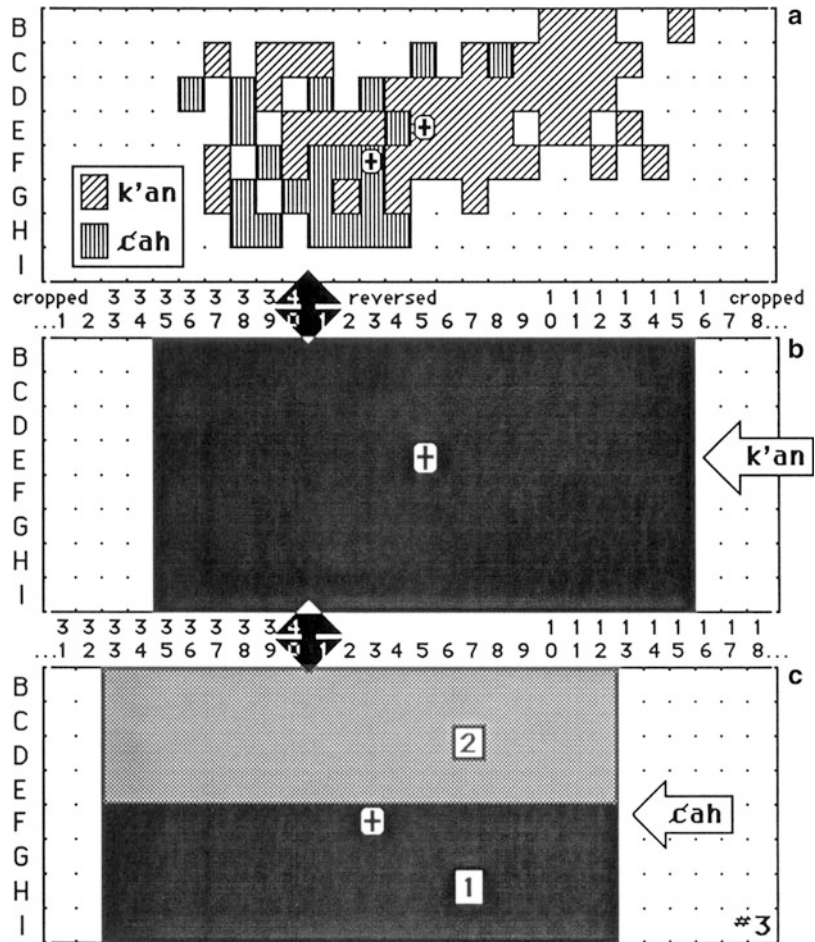
Figure 5 shows coextension in an early phase (closer to near synonymy than to inclusion).

Another example is the Zulu cool category in Fig. 4.

Coextension cannot be explained solely in terms of perceptual dimensions; instead, it is the observer who “assumes opposite slants on the same sensations and names them differently from each angle” [8, p. 113]. Coextension can thus be thought of as variable semantic bias or a strategy. This provides a strong argument in favor of subjectivity and speaker agency in categorization and meaning construction.

In inclusion, the naming and/or mapping ranges of the subordinate (recessive) term fall inside that of the superordinate (dominant) term.

Vantage Theory of Color, Fig. 5 Coextension in an early phase. Warm in Tzeltal (Mayan, Tzeltalan), Paraje Nabil, Tenejapa, Chiapas, Mexico, male 65, 1980, (a) naming and foci, (b–c) mappings. (Figure received from Robert E. MacLaury)



This happens when, as a result of strong attention to *D*, one of the ranges tends to “drift away” but both still share fixed cognitive coordinates. An example is the warm category in a speaker of Aguacatec (Awakateco), a Mayan language of Guatemala ([8], pp. 195–196, cf. also 10).

When the strength of *D* rises even more, the cognitive link between the vantages is broken, and the two ranges separate: inclusion becomes *complementation*, a relation between the dominant vantages of distinct categories. Extreme value of *D* causes category split.

Finer Distinctions

VT models finer aspects of color categorization as *frames*, *stress*, and *viewpoints*.

A frame is a closed system of interdependent parameters, such as emphases on *S* vs. *D*.

Occasionally, three vantages on a category are constructed, e.g., in the warm category in Aguacatec [8, pp. 115–116]. In a non-framed analysis, the three vantages are called *dominant*, *recessive*, and *ultra-recessive*, but a framed analysis links the dominant and the recessive terms (the red-focused *k'aq* and the yellow-focused *q'an*) into frame I, while the recessive and the ultra-recessive terms (*q'an* and the brown-focused *sq'inko'x*) constitute frame II. In frame II, the relationship between the recessive and the ultra-recessive term is analogous to that between the dominant and the recessive term in frame I. Thus, the terms *dominant*, *recessive*, and *ultra-recessive* are relative rather than absolute.

Stress is mental proximity to either the fixed or the mobile coordinates in a vantage. In the color domain, stress may be put on hues or on relations

between them. In the former case, one of the hues may “draw” the category to itself – this is called *skewing* [7]. The category may thus divide, as happened in the Mayan language of K’ekchi (Guatemala, Belize and El Salvador) [11, pp. 128–129]. In the latter case, focus placement may be random, foci may substitute for each other, and mismatches may be observed between naming ranges, foci, and mappings both for a single speaker and between informants. Informants may maintain the dominant-recessive pattern without preference for which hue is dominant or recessive: the pattern itself seems to be more important than the hues on which it is based. Stress on mobile coordinates has been noted in Mazatec (or Huautla, Oto-Manguan, Mexico) [8, 11] or Mam (Mayan, Guatemala) [8, pp. 294–306]. The behavior is consistent and appears chaotic only for speakers used to stressing fixed coordinates, as those of English.

Viewpoints are gradable degrees of subjectivity/objectivity with which a person constructs a category or conceptualizes an object or scene [8, 10, 11]. The notion has been found useful in modeling several Mesoamerican languages (Nahuatl, Cakchiquel, Northern Tepehuán, Quiché, Chinantec, or Lacandón). It is a good example of how VT can be extended beyond color: the conception has been applied in several accounts of linguistic behavior [4, 5, 8].

Other Notions

Although originating in the color domain, VT is claimed to have universal application and pertain to categorization at large. Aspects of VT not discussed here include, among others, brightness-based categories; full vs. partial inversion of coordinates; submerged vs. reflective vantages; the role of *S* and *D* in categorical evolution; the flip-flop, or oscillation, of color term meanings; dual and triple foci; non-discriminatory vs. analytic vs. synthetic thinking; the spotlight effect; and more [6, 8, 10]. On a more general level, VT relates to the question of linguistic relativity, representing a non-Whorfian stance but one which endorses variety in linguistic behavior.

The variety, however, occurs within a cognitively universal but plastic mechanism of categorization [9].

Vantage Theory’s Unique Contributions

With VT, MacLaury has contributed to research on color categorization in three major ways. First, he has redefined the notion of a category, traditionally identified with a color term. In VT a term names a vantage on a category, and there may be one, two, or occasionally three such vantages. Second, he has identified, defined, and modeled coextension as a relation between vantages that at face value appears chaotic or indicative of errors in data elicitation. Third, VT stresses individual variation and personal preference in color categorization. It is an attempt to model that variation and reconcile it with both the classical perceptual constructs of hue, brightness, and saturation and cognitive emphases on similarity and difference. Vantages on categories help to systematically explain how the same stimuli may be differently interpreted by different individuals but also how the same individual might interpret the same stimuli as different on different occasions.

Future Directions

So far, with few exceptions, VT has remained largely dormant though full of potential. In order to utilize this potential, it is proposed that the theory should be made more accessible to the color research community by producing textbook-type publications (a noteworthy precedent is Sect. “9.3” of [2], though it remains isolated and rather limited in scope). It may and should provide more insight into the linguistics of color and color naming, as indeed it has in the past [3, 12]. Also, VT as a theory is open to extension and elaboration.

Cross-References

- [Berlin and Kay Theory](#)
- [World Color Survey](#)

References

1. Allan, K.: Categorizing percepts: vantage theory. In: Brown, K. (ed.) *Encyclopedia of Language and Linguistics*, 2nd edn, pp. 252–253. Elsevier, Oxford (2005)
2. Allan, K.: *Natural Language Semantics*. Blackwell, Oxford (2001)
3. Anishchanka, A.: Vantage construal in the attributive use of basic color terms: the CAN and N of NC constructions. *Lang. Sci.* **32**(2), 170–183 (2010)
4. Glaz, A.: Let me hear you talk and I'll tell you where you are. *Naukovi zapysky Nizhynskogo Derzhavnogo Universytetu imeni Mykoly Gogola. Filologichni nauky* (Research Papers of Nizhyn University. Philology), pp. 90–97 (2007a)
5. Glaz, A.: Vantage theory: a newcomer to the cognitivist scene? In: Fabiszak, M. (ed.) *Language and Meaning*, pp. 91–112. Peter Lang, Frankfurt am Main (2007b)
6. MacLaury, R.E.: Vantage theory. In: Taylor, J.R., MacLaury, R.E. (eds.) *Language and the Cognitive Construal of the World*, pp. 231–276. Mouton de Gruyter, Berlin (1995)
7. MacLaury, R.E.: Skewing and darkening: dynamics of the cool category. In: Clyde, L.H., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 261–282. Cambridge University Press, New York (1997)
8. MacLaury, R.E.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, Austin (1997/2011)
9. MacLaury, R.E.: Linguistic relativity and the plasticity of categorization. In: Pütz, M., Verspoor, M. (eds.) *Explorations in Linguistic Relativity*, pp. 251–293. John Benjamins, Amsterdam (2000)
10. MacLaury, R.E.: Introducing vantage theory. *Lang. Sci.* **24**(5–6), 493–536 (2002)
11. MacLaury, R.E.: Vantage theory in outline. In: Glaz, A., Moist, M.L., Tribushinina, E. (eds.) *Vantage Theory: a View on Language, Cognition and Categorization*, pp. 66–136. Cambridge Scholars Publishing, Newcastle upon Tyne (2013)
12. Stanulewicz, D.: Polish terms for 'blue' in view of vantage theory. *Lang. Sci.* **32**(2), 184–195 (2010)

Variance Adaptation

► Color Contrast Adaptation

VEP

► Visual Evoked Potentials

Visible Light Communication

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Synonyms

VLC

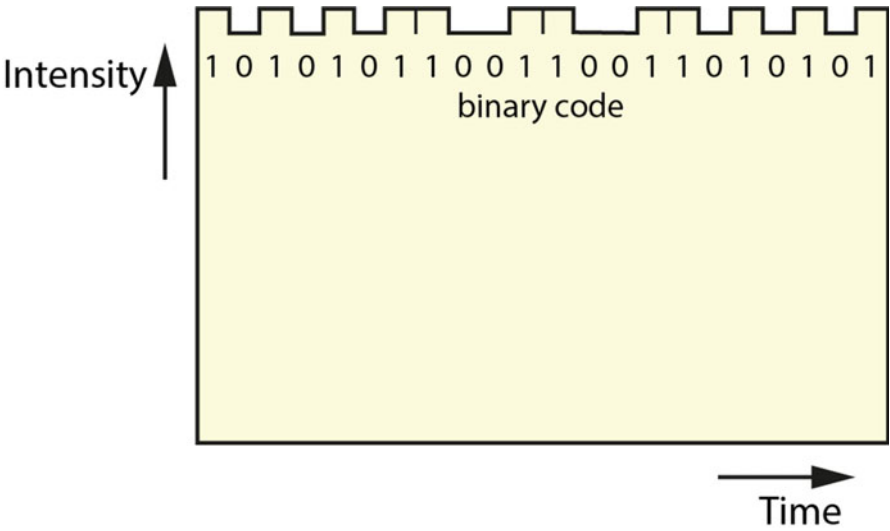
Definition

Light in free space from LED luminaires used simultaneously for illuminating the space and wirelessly transferring data into that space.

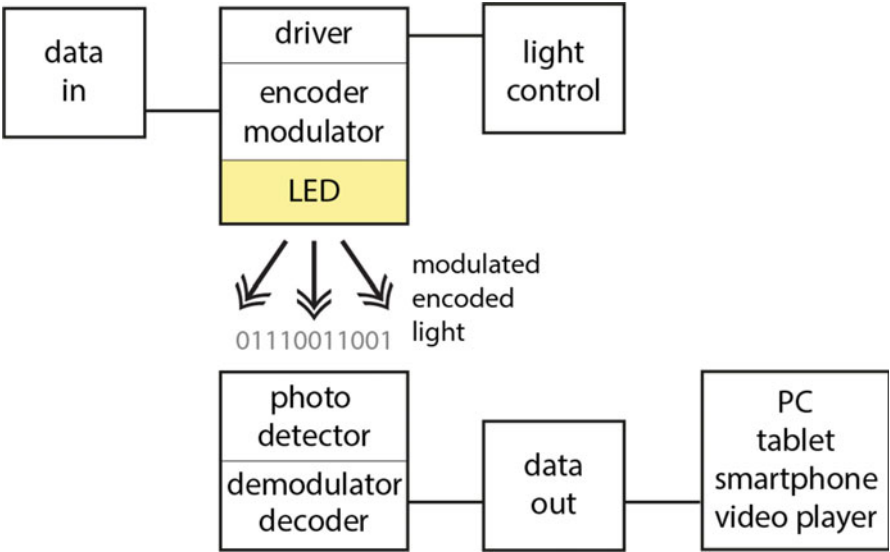
Principle of Visible Light Communication

In a visible light communication (VLC) lighting system, data can be encoded by switching the system's LED luminaires on and off at high frequency, i.e., modulating the intensity of light. In the VLC practice, the light is not switched on and off entirely but dimmed and not dimmed. So, only the top part of the intensity is modulated, resulting in only a small ripple in intensity (Fig. 1). In this way, the illumination level is hardly affected by the modulation. Figure 2 shows the functional scheme of a VLC system. It exists of a LED luminaire with a modulator that codes the light and a photodetector, somewhere in the lighted space, with a demodulator to decode the data. The photodetector can, for example, be a simple camera of a smartphone. A smartphone app can do the demodulation. For PCs, laptops, and other smart apparatus, low-cost dongles are available with an integrated photodetector and a demodulation/decoder processor.

Of all light sources, only LEDs are suitable for visible light communication. This is because the luminous flux of LEDs varies nearly instantaneously with a variation of current through the LED. Only in this way, light can be modulated at high speed, i.e., the light “on” and light “off” (or “not-dimmed” and “dimmed”) condition can



Visible Light Communication, Fig. 1 Example of encoding by modulating the top of the luminous intensity of a LED luminaire



Visible Light Communication, Fig. 2 Visible light communication (VLC) functional scheme

be produced at extremely high frequencies. The response time using incandescent and gas discharge lamps is too long. With phosphor LEDs, the response time is somewhat longer than with non-phosphor LEDs, but still short enough for VLC with moderate to high transfer rates (up to 100 Mbit/sec). The phosphor-converted components of the light of

a white (blue-pumped) LED can be filtered with a narrowband blue filter placed in front of the photo-detector. In this way, only the light of the blue LED, with near-zero response time, can be used for VLC. In this way, white phosphor LEDs can have the same ultrahigh transfer rates as non-phosphor LEDs (up to more than 1 Gbit/sec). So, commercial,

off-the-shelf LED luminaires can be used for visible light communication.

To ensure that the modulation does not interfere with the illumination quality and to permit for data communication under all dimming conditions, more sophisticated modulation techniques are often used. Light reaches the photodetector not only directly but also indirectly. It makes the demodulation technique even more complicated. The method must permit for making a distinction between the direct and indirect light components. Such a method also helps to solve the problem of direct light sometimes being intercepted by objects or persons in the space, resulting in the photodetector having no longer a direct view of the light source. In such a situation, with appropriate modulation techniques, the reflected light can keep the communication going on. Interference by daylight and other ambient light sources also puts requirements on the modulation and demodulation technique.

As light does not penetrate through opaque walls and thus all communication is contained within the room, the security of VLC is higher than with RF communication.

Download-Only Systems

Many VLC applications concern the use of VLC as a tool for pushing information, i.e., downloading only or broadcasting. Reasons for using VLC instead of Wi-Fi (radio frequency) are no congestion as with the Wi-Fi bands, lower installation cost, lower energy consumption, higher security, no electromagnetic interference, and higher data transfer speeds. New applications using data pushing have emerged and will continue to emerge for situations where the installation of RF systems requires too much effort. Some examples are described below.

Personal Movie System in Planes

The LED reading lamp in planes can be used to transfer sound and video for the in-flight personal sound and movie system. Since dedicated communication cables are no longer needed, the robustness of the system increases considerably, while the fuel consumption of the plane decreases because of the eliminated weight of cables.

Indoor Navigation

A GPS (global positioning system) uses the signals received from satellites at different positions to determine the location of the GPS. The satellite signals are not strong enough to penetrate buildings. Consequently, GPS can only be used outdoors. With a VLC system, it is possible to make the LED luminaires used for the lighting in a building, acting similar to the satellites in a GPS. Such a system is either called a visible light positioning (VLP) system or an indoor positioning system (IPS). The satellite signals received by a GPS are used to calculate the distance of the GPS from the known locations of the satellites. In an indoor VLC navigation system, data transferred by the luminaires contain an identification of each luminaire. With this information, the location of each luminaire can be determined from that data so that, in turn, the position of the camera can be calculated.

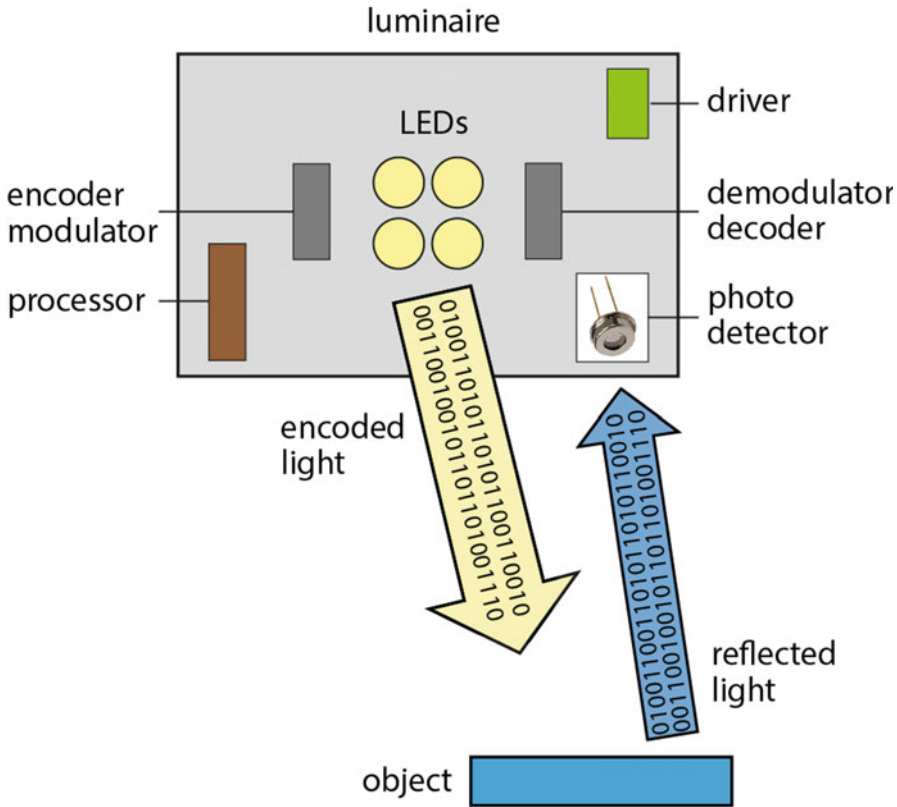
Light Itself as Sensor

By adding to the luminaires of a VLC system photodetector, it is possible to measure the time of flight, ToF, of light from the moment it leaves a luminaire until its small, reflected component returns at the luminaire from which it originated (Fig. 3). In this way the VLC system can be used for scanning a room to pinpoint the location of objects in an area: light as a sensor, also referred to as time-of-flight, ToF, sensor.

The contours of objects and even the pose and movements of persons (sitting, standing, lying, or walking) can be determined with time-of-flight sensors [1, 5, 7]. Sudden changes in the pose from standing to lying can be recognized as a person who has fallen. It can subsequently set an alarm. This emerging feature of light as a sensor, once technically perfected, would be well suited for use in healthcare environments and homes of (older) singles living alone. Many other applications can be thought of.

Combined Down-Upload Systems (Li-Fi)

VLC can be extended into a bidirectional communication system with a down- and uplink. Harald Haas of the University of Edinburgh introduced at



Visible Light Communication, Fig. 3 Light as a sensor. The time of flight of encoded light of a luminaire leaving the luminaire until its reflected light returns at the luminaire is used to determine the location of an object in a room

TEDGlobal in 2011 the name Li-Fi (light fidelity) for such bidirectional systems [3]. It is a much-needed alternative for, or a complement to, the more and more congested Wi-Fi wireless RF communication system. Of course, it would be disturbing if for the uplink visible light is used. Disturbing light beams would be radiating from all connected devices in the room, such as PCs, laptops, and smartphones. So, while with Li-Fi the downlink indeed uses visible light, the uplink uses either invisible infrared or Wi-Fi. Dongles for use in PCs and laptops with an integrated photodetector and infrared or Wi-Fi source have been introduced. A room connector must be used to interconnect Li-Fi networks in adjacent rooms. It sends data from one side of a wall to the other side through an optical fiber. In most data communication situations, much more data is downloaded than uploaded. Therefore, even by using a

Li-Fi system with the upload link using Wi-Fi, still, a substantial contribution to reducing the congestion of Wi-Fi networks is obtained [2]. Because of the explosive use of the Internet of things, the limits of the capacity of Wi-Fi will be reached quickly. It makes a fast roll-out of Li-Fi, with its virtually infinite bandwidth, indispensable. It may be expected that this fact, together with the energy efficiency, security, and ease of installation of Li-Fi systems, will be a strong drive for the fast development of the global Li-Fi market.

The Li-Fi market will certainly not only consist of lighting installations with the dual function of illumination and data transfer. Dedicated communication-only installations using laser LED beams (in the infrared wavelength region) are also being developed where transfer speeds of 100 Gbits/sec are foreseen [4, 6].

Cross-Reference

► Internet of Things and LEDs

Acknowledgments Adapted by permission from Springer Nature “[Interior Lighting, fundamentals, technology and application](#)” by Wout van Bommel (2019).

References

1. Bhattacharya I, Radke RJ (2016) Arrays of single pixel time-of-flight sensors for privacy preserving tracking and coarse pose estimation. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) 1–9
2. Dimitrov S, Haas H (2015) Principle of LED light communication; towards networked Li-Fi. Chapter 4: digital modulation-schemes. Cambridge University Press, UK
3. Haas H (2011) Wireless Data from Every Light Bulb, TEDGlobal, Aug 2011 <http://bit.ly/tedvlc>. Accessed on 1 November 2018
4. Haas, H.: LiFi is a paradigm-shifting 5G technology. Review in Physics. 3, 26–31 (2018)
5. Karlicek RF, Radke RJ, Little TDC, Butala PM, Jia L (2016) Sensory lighting system and method for characterizing an illumination space. US Patent nr 9363859
6. Koonen AMJ, Oh CW, Mekonnen K, Tangdiongga E (2015) Ultra-high capacity indoor optical wireless communication using steered pencil beams. Proceedings of the International Topical Meeting on Microwave Photonics (MWP2015), Paphos, Cyprus
7. Woodstock T-K, Radke RJ, Sanderson AC (2016) Sensor fusion for occupancy detection and activity recognition using time-of-flight sensors. 2016 19th International Conference on Information Fusion (FUSION) 1695–1701

Vision and Age

Ken Sagawa

Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Synonyms

[Age-related changes of human seeing abilities](#)

Definition

Aging of the human eye and its influence on seeing abilities with regard to brightness, color, spatial resolution (visual acuity and legibility), visual field, and glare.

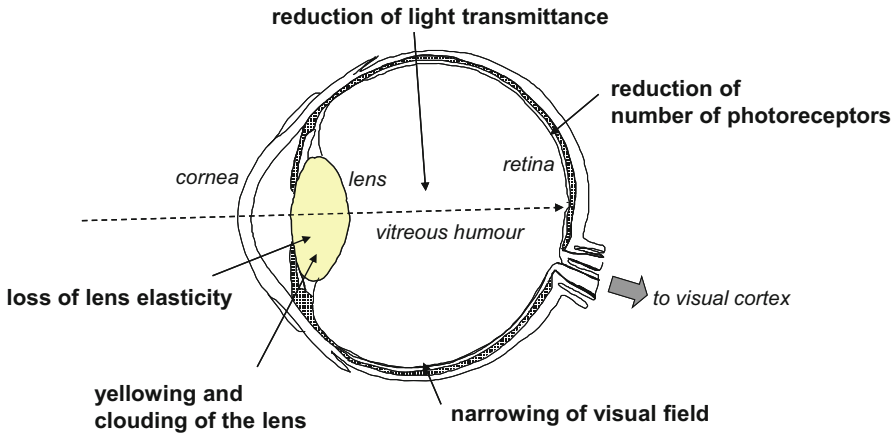
Senescent Changes in the Visual Organ

Visual information accounts for about 70% or 80% of sensory information in human beings. Like other sensory organs, human vision has senescent changes for most of the parts in the visual system, and these make quality of seeing degraded to older people.

Figure 1 shows a simple sketch of human eye together with major structural changes with age that affect perception of vision. Light from the external world enters into the eye through the lens and reaches onto the retina that covers the back of the eyeball. Two age-related changes occur at the lens; one is loss of elasticity and the other one yellowing and clouding of the lens. The former one brings on less accommodative power of the lens and consequently results in loss of fine-tuning of visual image, i.e., visual acuity, especially at near point less than 1 m of viewing distance. The latter one makes color appearance yellowish and hazed and also makes visual acuity worse due to degraded retinal image by scattered lights. Reduction of light transmittance also occurs in vitreous humor in the eyeball by absorption and scattering of light which also reduces performance for most of the visual functions.

In the retina, photoreceptors that are densely packed in the retinal layer decrease their density with age, and this brings the reduction of visual acuity too and furthermore, particularly in the peripheral retina, less sensitivity to temporal changes of lights, such as flickering or flashing lights. The decrease of photoreceptors in peripheral retina also results in narrowing visual field and makes the detection of lights presented in periphery more difficult for older people.

Visual responses generated in the retina are finally transferred to visual cortex in the central brain. Some senescent changes also occur in this



Vision and Age, Fig. 1 An illustration and major senescent changes of the eye

pathway of neural transmission, and consequently the quality of visual responses is degraded more or less.

Senescent changes of human visual system and resulted performance changes for older people were summarized in some scientific literature [1–3] and in documents for the application fields of lighting and visual signing [4, 5].

Following are major functional changes of vision caused by the senescent changes of the eye described above.

Spectral Sensitivity of the Eye and Luminance

Spectral Sensitivity

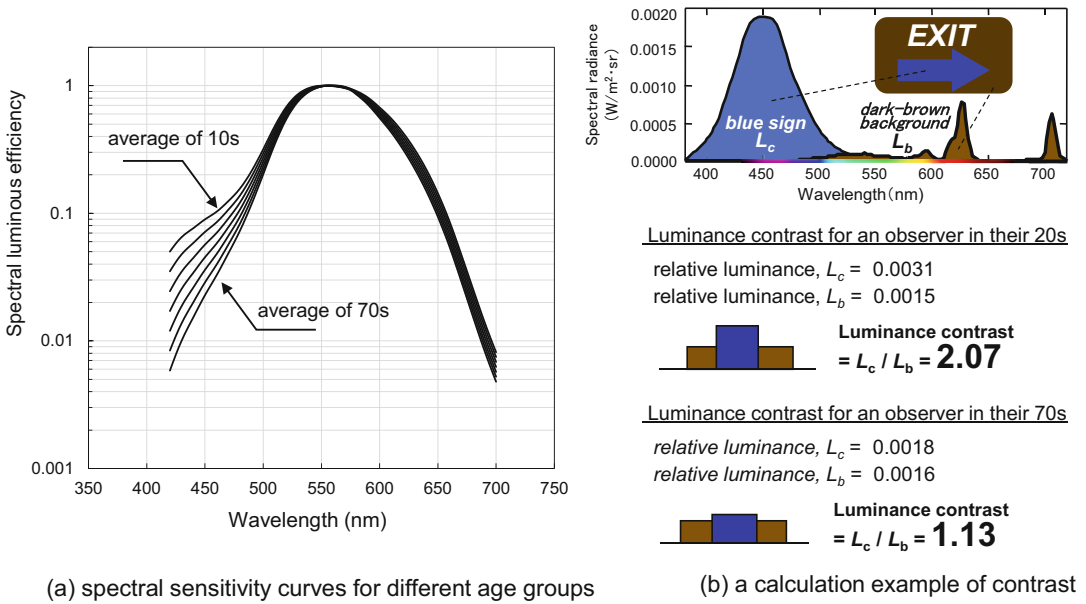
Human visual system responds to electromagnetic radiation ranged in wavelength about 400–700 nm called visible range (correctly defined as 380–780 nm by CIE). Most sensitive part is at around 555 nm (green-yellow), and the sensitivity decreases gradually toward both the shorter and the longer end of the visible range. Out of this range, i.e., the ultraviolet range (shorter than 380 nm) and the infrared range (longer than 780 nm), the eye does no longer respond to and cannot see the radiation. The general shape of this sensitivity curve along with the wavelength is a bell-shaped smooth curve with a peak at 555 nm.

This spectral sensitivity of the eye changes with age. Due to the yellowing of the lens mentioned in the previous section, spectral lights in the shortwave region (about 400–500 nm colored violet or blue) are partly absorbed by the lens and do not fully reach to the retina, and consequently the shortwave lights become less sensitive to the eye. The absorption gradually increases with age, and the violet or blue lights look darker to older people [6, 7].

Figure 2a shows the spectral sensitivity curves of the eye for seven different age groups from teens to 70s in decade steps all being normalized to 1.0 at its peak sensitivity [7]. It is clear that sensitivity at the shortwave region gradually decreases with age, and the difference reaches to more than one logarithmic unit between teens and 70s at the short end of the spectrum. This practically means brightness for a blue lamp or a blue LED display that generates the shorter wavelengths looks much darker to older people compared with young people.

Luminance Contrast for Older People

Spectral sensitivity is a fundamental function to estimate brightness of a light or an object which is technically called luminance in the field of photometry, i.e., the visually meaningful metric of light. This luminance, denoted as L , is defined by CIE as follows:



Vision and Age, Fig. 2 (a, b) Spectral sensitivity of human eye for seven age groups from teens to 70s [7] and a calculation example of contrast for a blue sign presented on a dark-brown background [8]

$$L = K_m \int L_{e,\lambda} V(\lambda) d\lambda \quad (1)$$

where

K_m maximum efficacy (683 lumen/watt)
 $L_{e,\lambda}$ spectral radiance (watt/m²·sr)
 $V(\lambda)$ CIE standard luminous efficiency (spectral sensitivity of the eye)

As expressed in this equation, luminance is a total amount of radiation weighted by visual sensitivity $V(\lambda)$. However, as $V(\lambda)$ changes with age (see Fig. 2a), the luminance changes its value accordingly for people with different ages.

Figure 2b shows an example of the calculation. Given that spectral radiation for a blue sign and a dark-brown background as shown at the top of Fig. 2b, relative luminances of them for observers in their 20s and 70s, for example, are calculable using sensitivity curves of 20s and 70s, respectively (the second curve from the top and the bottom one in Fig. 2a), and luminance contrast between the blue sign and the background is obtained. The results are 2.07 and 1.13 for the young and older observer respectively, which

means that the same sign is seen in good contrast for the young person but in lower contrast for the older person. With this way of calculation, contrast of any colored sign or object can be evaluated accurately. This is called “age-related luminance contrast.” To avoid any confusion with the already standardized luminance by CIE using $V(\lambda)$ (the standard luminous efficiency curve), this calculation only applies to know the contrast, i.e., relative relationship of luminance [8].

Color Perception

Color Appearance

Due to the yellowing of the eye lens with age, all the perceived colors look more or less yellowish for older people. This color shift is not so vivid nor conscious due to the adaptation or constancy of the visual system. In some case, however, when we try to discriminate a yellow sign on a white background or vice versa, for example, it becomes harder for older people to discriminate them.

Scattering light inside the eye (lens and vitreous humor) also changes color appearance. Scattered lights falling on the retina make all the colors

desaturated like white noise and consequently reduce the vividness of color (saturation). Due to this desaturation effect and together with other neural pathway changes with age, the discrimination of color becomes worse with age in a whole perceived color space [1].

Color Category

Human color space consists of a large variety of colors that distribute in a three-dimensional space of hue, lightness, and chroma or saturation. Recent psychological studies on color perception have found that these colors are perceived into a limited number of groups in higher central brain, namely, color category. For example, a purplish red and an orangish red are both perceived as a family of red group regardless of the small difference. This concept can be usefully used for color combination in a manner that any two-color combination made from different categories becomes differential or, on the contrary, any two colors picked up from the same category are easily confused. Data on which color belongs to which category and how the aging affects the category is a critical issue for using this concept and application.

Figure 3a shows the data for color category for young and older people expressed in Munsell Color Space measured on the bases of subjective evaluation of color similarity for a total of 13 fundamental colors as a focal color of each category [9]. Most important and interesting point of the color category data is the fact that span of the category is smaller for older people than that for young people for almost all the fundamental colors. This is a feature contrary to color discrimination. One of the reasons for this might be the desaturation of colors for older people caused by interocular scattering of light. Vivid colors reduce their saturation and shrink their areas in the color space. Figure 3b is an example of how to apply the color category data. Coloring of six traffic lines is taken as an example. The six lines are colored with six different colors each of which are picked up from color categories that do not overlap each other in the color space. This makes the color combination discriminable to older people if the category data of older people (solid lines in Figure 3b) apply. Any color belonging to the

span of category is allowed to use for this color combination with no need for picking up exact colors predetermined, which is a practical advantage for creating color combination.

Contrast Sensitivity

Contrast sensitivity is a visual ability to perceive a difference between any two adjacent fields either in luminance or in color, or both, such as an edge between a wall and a door, marking on steps, characters or symbols on an extended background, etc. Visibility of signs and displays depends on the contrast of their contents, and particular care or attention is required for a better contrast.

As shown in Fig. 4, contrast sensitivity of human visual system is usually specified with a periodic pattern like a sine- or a rectangular-wave pattern of variable number of cycles per degree (cpd) with luminance modulation between a darker ($L_{\min}$) and a brighter ($L_{\max}$) part which are used as a metric of contrast such as Michelson contrast.

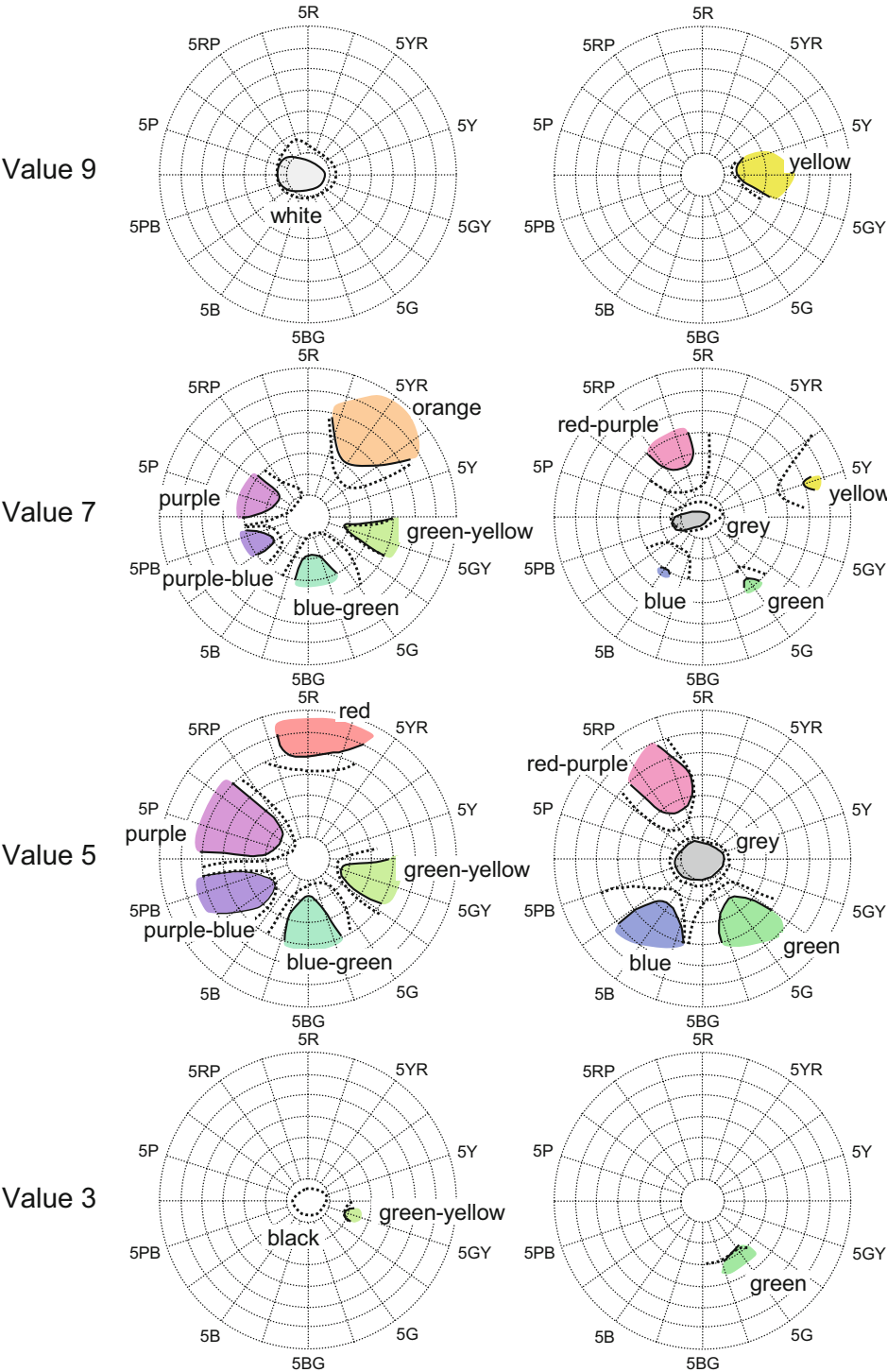
The right side figure is one example of contrast threshold data, i.e., minimum contrast to detect a grating pattern. The contrast threshold is minimum (i.e., most sensitive) at around 2 or 3 cpd and increases toward the higher and the lower spatial frequency region. Age-related difference is seen at the higher-frequency region meaning that older people need higher contrast to see a spatially fine or detailed pattern.

Some studies on contrast sensitivity function (CSF), which is a reciprocal of contrast threshold as a function of spatial frequency, have revealed there is a clear age-related decline of sensitivity with age at higher-frequency region as demonstrated in Fig. 4, which are summarized in CIE Report [4].

Visual Acuity and Legibility of Characters

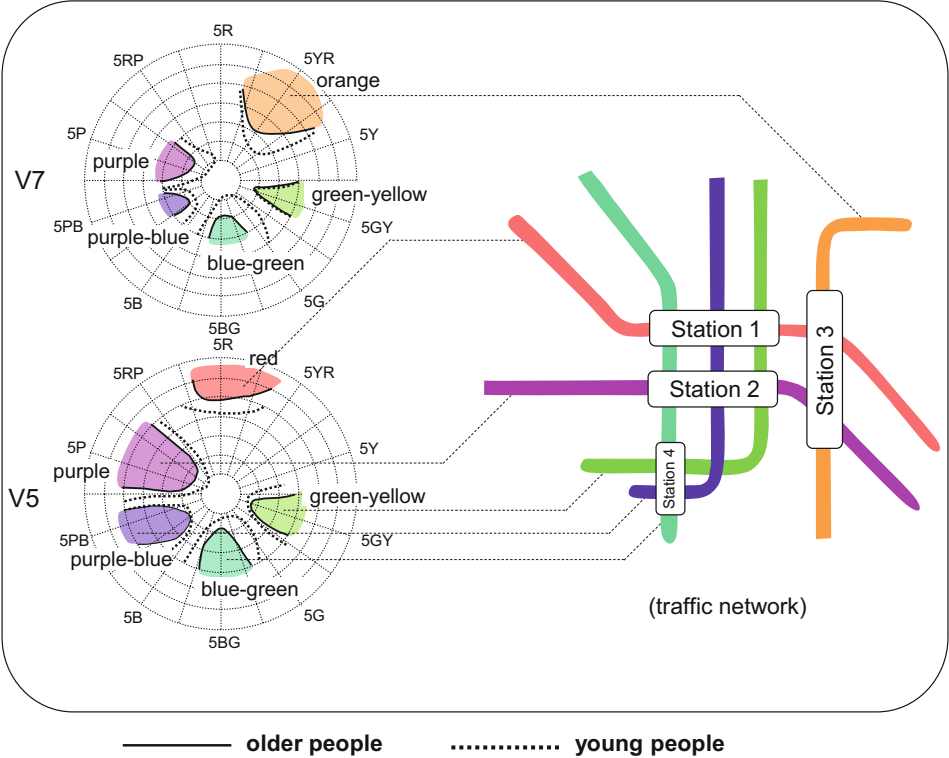
Visual Acuity

Visual acuity is an extent to which human visual system can resolve spatial details of visual stimuli.



(a) Color category data

Vision and Age, Fig. 3 (a, b) Spans of color category for fundamental colors and an application example. (Redrawn from Ref. [9] and ISO24505:2016 [10]. Coloring is image only)



Vision and Age, Fig. 3 (continued)

It is defined quantitatively, in case of decimal acuity, as a reciprocal of minimum visual angle expressed in minutes to detect a gap of a circle (called Landolt C) as follows:

$$V = 1/\Delta\theta \tag{2}$$

where

- V visual acuity
- $\Delta\theta$ minimum detectable visual angle in minutes of arc

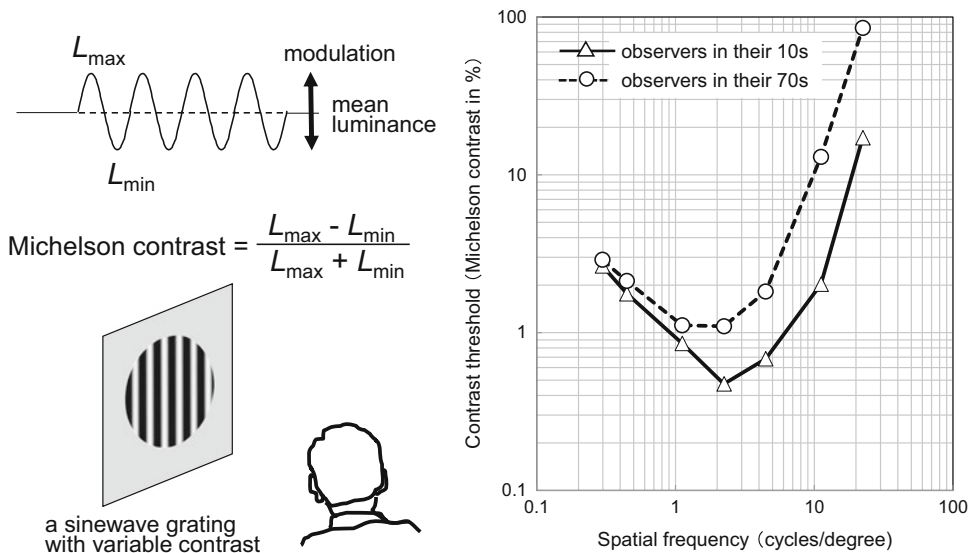
There are three major critical factors that affect visual acuity: age, viewing distance, and luminance. Figure 5a shows how visual acuity changes with viewing distance, and Fig. 5b demonstrates luminance effect, both for different age groups from teens to 70s [12]. It is clear from Fig. 5a

that the visual acuity at near distance less than 1 m remarkably decreases for age groups over 40 years indicating that reading out fine-printed information such as maps or newspapers at hand is much harder for older people.

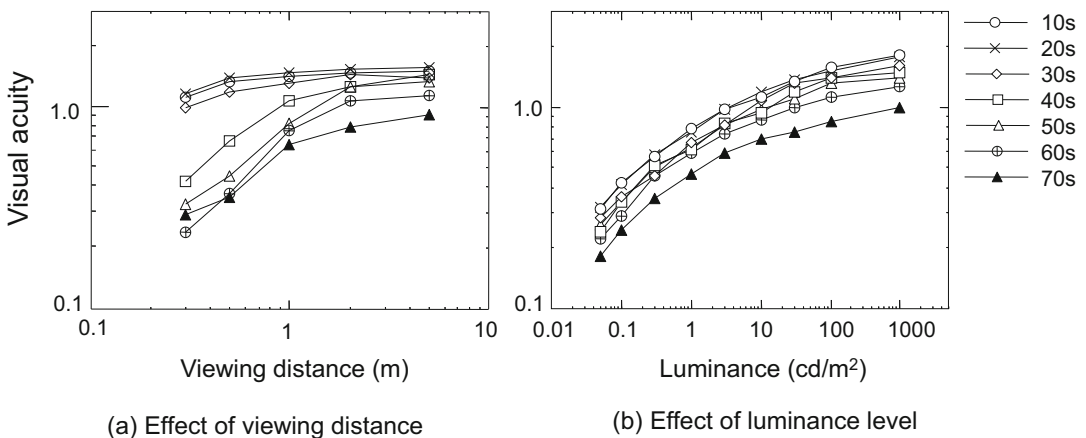
With regard to luminance effect (Fig. 5b), visual acuity becomes gradually worse as luminance decreases which shows discrimination of visual signs at a darker condition becomes difficult for both young and older people. It should be noted that the luminance change affects all the age groups in the same manner and no remarkable difference in the relative course of luminance effect is found among the age groups.

Legibility of Characters

A visual task that is most related to visual acuity is identification of characters or reading sentences, i.e., legibility or readability, where discrimination



Vision and Age, Fig. 4 Definition of contrast and an example of the contrast threshold data for observers in their 10s and 70s. (Data redrawn from Ref. [11])

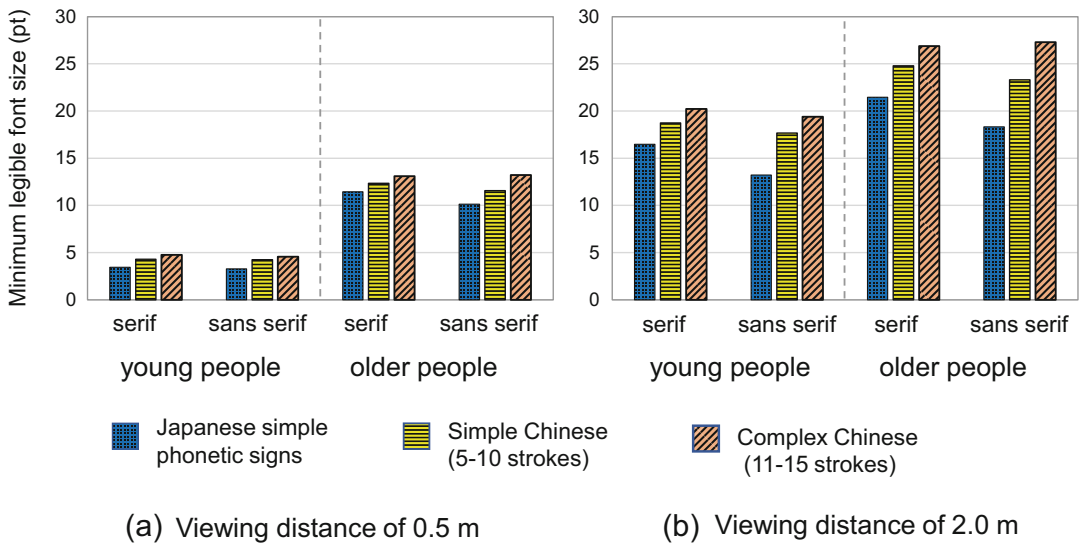


Vision and Age, Fig. 5 (a, b) Visual acuity as a function of (a) viewing distance and (b) luminance for seven different age groups [12]

of fine visual pattern is required. Smaller font size is sufficient if the visual acuity is higher and, on the contrary, larger font is required for lower visual acuity.

Figure 6a, b are examples of how much font size is required to see at moderately bright level (100 cd/m²) for Japanese characters of phonetic signs (including Arabic numerals) and simple and complex Chinese characters [12]. The data show

the minimum legible font size, under which the characters are not legible, for young and older people at 0.5 and 2.0 m of viewing distance. At the short distance, minimum legible font size for young people is about 4 points in average, but at the same condition, older people require about 12 points, i.e., three times larger than that for young people. This age-related difference becomes smaller at the longer viewing distance



Vision and Age, Fig. 6 (a, b) Minimum legible font size for Japanese characters. (Redrawn from Sagawa et al. [12])

of 2 m (Fig. 2b). It is also clear from these data that the size for the sans-serif font is slightly but consistently smaller than that for the serif font for almost all the conditions indicating that the sans-serif font is more legible than the serif font when the viewing condition is identical.

Practically it is useful to estimate the minimum legible font size at any viewing condition where age, viewing distance, and luminance are varied. An estimation method for the minimum legible font size has been provided by ISO [13]. Some calculated examples at a constant luminance condition, for example, at the 100 cd/m² level, are shown in Fig. 7a, b as a function of viewing distance for different age groups.

Visual Field

Human vision covers a wide range of visible field, not only a center field around visual axis but also a peripheral field that extends to almost 180° in front of the eye. This visible range is called visual field, the size of which is usually expressed in visual angle in degree from the central visual axis regardless of viewing distance. The visual field is functional, and the size of which depends on what we see or what we do with vision. This is defined as

useful field of view or useful visual field. The useful visual field for simple detection of a light, for example, is rather wide, but the field for complex visual task like reading letters becomes much smaller being limited to the central part of the visual field.

The size of useful visual field is affected by age. Due to the decrease of density of photoreceptors at peripheral retina with age, the visual field size is reduced for older people compared with young people. Figure 8 shows the experimental data on useful visual field for detecting a target with different luminance contrasts, target colors, and adaptation levels [14]. Visual field size reduces with reduction of target contrast, depends on target color, and reduces with adaptation level from the photopic to mesopic level, and it is clear that older people consistently show smaller visual field than young people for all the viewing conditions. Traffic signs or emergency signs could be missed to see by older people if they are presented to the peripheral visual field.

Characteristics of visual field in ordinary scene are much more affected by cognitive aspect of visual information such as attention, searching, memory and learning, etc. Understanding of visual cognition and its aging effect in relation to useful visual field, however, are not well-known yet and will require further studies in the future.

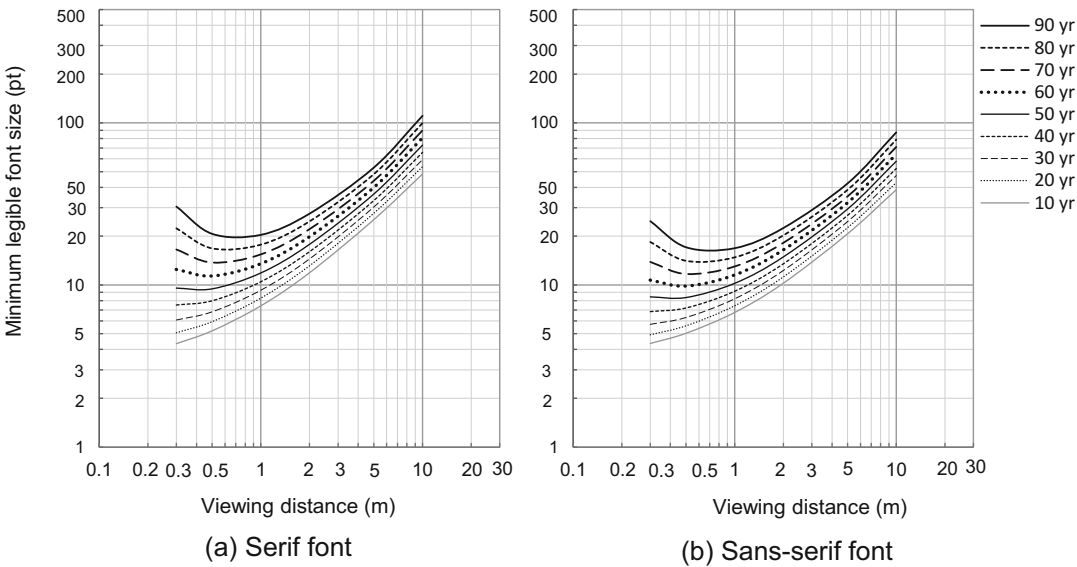


Figure 7 (a, b) Minimum legible font size at the luminance level of 100 cd/m² as a function of viewing distance for different age groups. (Redrawn from ISO 24509 [13])

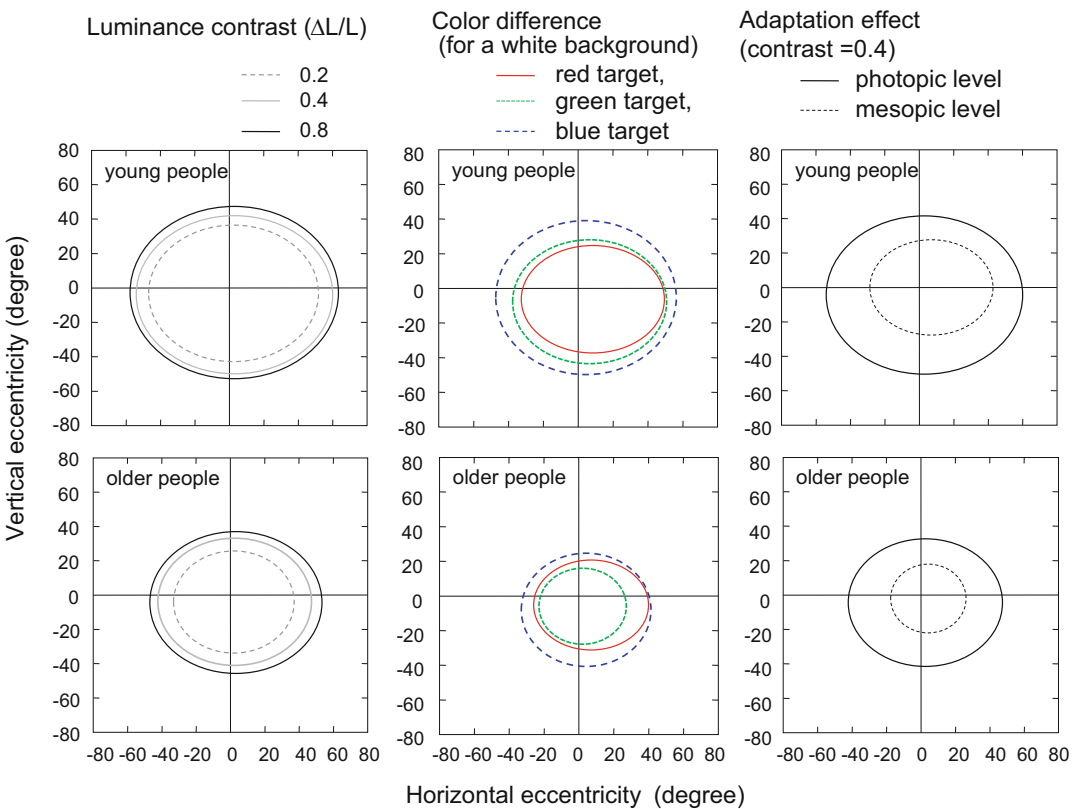


Figure 8 Useful visual field for detecting a target on a uniform gray background with variable contrast, color of the target, as well as adaptation levels. (Redrawn from the AIST data [14])

Glare

It is well-known that older people are more sensitive to glare than young people due to the light scattering in the ocular media that increases with age. Though it is obvious from previous studies that the physical light scattering in the ocular media of the eye increases with age [15], psychological data are scarce that show the age-related increase of subjective feeling of glare. Further studies and data collection to confirm the aging effect on psychological evaluation of glare are required.

Summary

Studies on vision and age have been conducted extensively over the past few decades, and basic knowledge has been cumulated to some extent, partly in response to the increasing social need for universal design or accessible design. The knowledge of aging effects on vision could be applied to the design of lighting, visual environment, and signage for better and safer life of older people. Further basic studies are emerging into a more challenging field of vision-based cognitive science and aging effects toward full understanding and application of human visual information processing.

Cross-References

- [Color Categorical Perception](#)
- [Color Combination](#)
- [Glare](#)
- [Spectral Luminous Efficiency](#)

References

1. Werner, J.S., Peterzell, D.H., Scheetz, A.J.: Light, vision, and aging. *Optom. Vis. Sci.* **67**, 214–229 (1990)
2. Owsley, C.: Aging and vision. *Vis. Res.* **51**, 1610–1622 (2011)
3. Haegerstrom-Portnoy, G.: The Glenn A. Fry Award Lecture 2003: vision in elders – summary of

- findings of the SKI study. *Optom. Vis. Sci.* **82**, 87–93 (2005)
4. Commission Internationale de l'Eclairage (CIE): CIE guide to increase accessibility in light and lighting, CIE 191 (2011)
5. International Organization for Standardization (ISO): Ergonomics data and guidelines for the application of ISO/IEC guide 71 to products and services to address the needs of older persons and persons with disabilities, ISO TR22411 (2008)
6. Kraft, J.M., Werner, J.S.: Spectral efficiency across the life span: flicker photometry and brightness matching. *J. Opt. Soc. Am.* **A11**, 1213–1221 (1994)
7. Sagawa, K., Takahashi, Y.: Spectral luminous efficiency as a function of age. *J. Opt. Soc. Am.* **A18**, 2659–2667 (2001)
8. International Organization for Standardization (ISO): Ergonomics – accessible design – specification of age-related luminance contrast for colored light, ISO 24502 (2010)
9. Sagawa, K., Takahashi, Y.: Span of categorical colours measured by similarity of colors. In: Proceedings of 25th Session of the CIE **1**, D1-64–D1-67 (2003)
10. International Organization for Standardization (ISO): Ergonomics – accessible design – method for creating colour combinations taking account of age-related changes in human colour vision, ISO 24505 (2016)
11. Ujike, H., Sagawa, K.: CSF as a function of age and its application to visibility of letters. In: Proceedings of CIE Symposium 2002 **CIE X25**, 81–86 (2002)
12. Sagawa, K., Ujike, H., Sasaki, T.: Legibility of Japanese characters and sentences as a function of age. In: Proceedings of the 15th Congress of the International Ergonomics Association **7**, 496–499 (2003)
13. International Organization for Standardization (ISO): Ergonomics – accessible design – a method for estimating minimum legible font size for people at any age, ISO 24509 (2019)
14. National Institute of Advanced Industrial Science and Technology (AIST): Database of sensory characteristics of older persons and persons with disabilities. <http://scdb.db.aist.go.jp/?lng=en>. Accessed 18 Dec 2019
15. Commission Internationale de l'Eclairage (CIE): CIE collection on glare, CIE 146/147 (2002)

Visual Appearance

- [Appearance](#)

Visual Contamination

- [Color Pollution](#)

Visual Evoked Potentials

Neil Parry

Vision Science Centre, Manchester Royal Eye Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
Centre for Ophthalmology and Vision Sciences, Institute of Human Development, University of Manchester, Manchester, UK

Synonyms

Event Related Potential; ERP; VEP

Definition

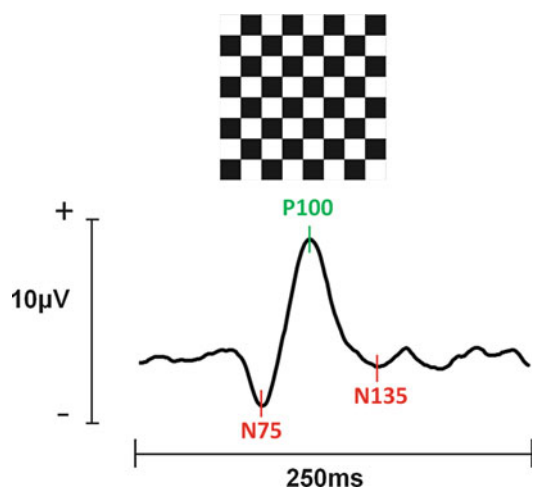
The visual evoked potential (VEP) is a means of extracting from the spontaneous electrical activity in the brain, electrical changes that are directly related to a specific brain action. It is a means of analyzing the Electroencephalogram (EEG). The event-related information ("signal") is extracted from the overall electrical activity ("noise") by some form of correlation. Usually, repeated samples of EEG are triggered by a particular event (e.g., an abrupt visual change) and averaged together. Because this event causes a particular type of brain response, at a predictable time, the process of averaging these time locked responses accentuates the event-related signal and smoothes out any uncorrelated activity. By using a stimulus that only contains time-locked changes in color, the VEP can provide an indirect measure of color information processing in the brain.

Scalp-Recorded Electrical Signals from the Brain

The brain is constantly generating large amounts of noisy electrical signals, emanating from the

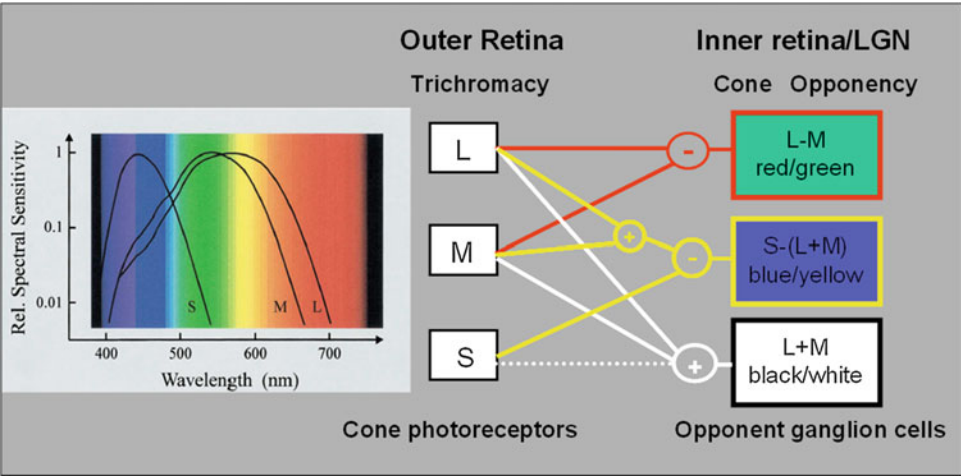
vast number of neurones firing at any given time. This ever-changing activity can be recorded by attaching electrodes to the scalp. The spontaneous activity, known as the Electroencephalogram (EEG) (<http://www.springerreference.com/docs/html/chapterdbid/116485.html>), reflects all the current motor, sensory and cognitive processes. If we wish to study brain activity that is related to a particular function (e.g., hearing or seeing) we can isolate that part of the EEG which occurs in response to specific stimulation by sounds or visual stimuli. In order to elicit such stimulus-related visual response, known as the visual evoked potential (VEP), it is necessary to cause a *change* in the stimulus, since the evoked potential is actually a measure of a *change* in electrical activity.

The crudest form of visual stimulus would be an instantaneous flash of light. This is rather like taking a sledgehammer to crack a nut, and is not very well suited to probing specific types of brain function. However it is often used clinically to get some idea of whether there is a gross connection between the eye and the brain, for instance in



Visual Evoked Potentials, Fig. 1 VEP elicited to a black and white checkerboard with each square measuring 1 degree of visual angle. The checks reversed every 250 ms. The key wavelets are negativities at about 75 and 135 ms, and a positivity at about 100 ms. The waveform was produced by averaging together fifty 250ms samples of EEG, each of which was triggered by an identical event (the instantaneous reversal of the checkerboard). Timings are relative to this trigger event

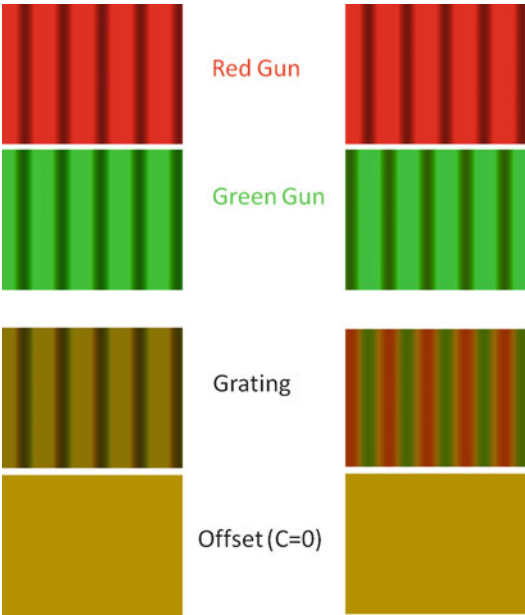
Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/978-3-030-89862-5_107.



Visual Evoked Potentials, Fig. 2 Spectral characteristics of the cone photoreceptors; subtractive and additive mechanisms giving rise to opponency; with thanks to Prof D. McKeefry

cases of cortical visual impairment. Patterned stimuli are more frequently employed; typically these are carefully specified repetitive patterns such as checkerboards or gratings. These have to be temporally modulated to elicit a response; usually this takes the form of pattern reversal or pattern onset-offset. Taking the example of a black-white checkerboard, in pattern reversal mode the black and white squares exchange places at fixed intervals, every few hundred milliseconds. In on-off presentation, the checks are replaced periodically by a blank gray screen so that there is no change in space-averaged color or luminance. It is crucial that artifacts are avoided: an overall luminance change between the two phases would itself elicit an evoked response, but this would have no relationship to the spatial content of the stimulus. Figure 1 shows a typical response to a reversing black and white checkerboard; this has a widespread use in clinical ophthalmology and neurology.

The electrogenesis of the evoked potential is not well understood, but the signals recorded from an electrode placed on the scalp above the occipital (visual) cortex are said to reflect mainly the activity of early visual areas V1 and possibly V2. The signals are recorded using a differential physiological amplifier, which takes the difference between the potential at the “active” site (e.g., the occiput) and a second electrode placed at a

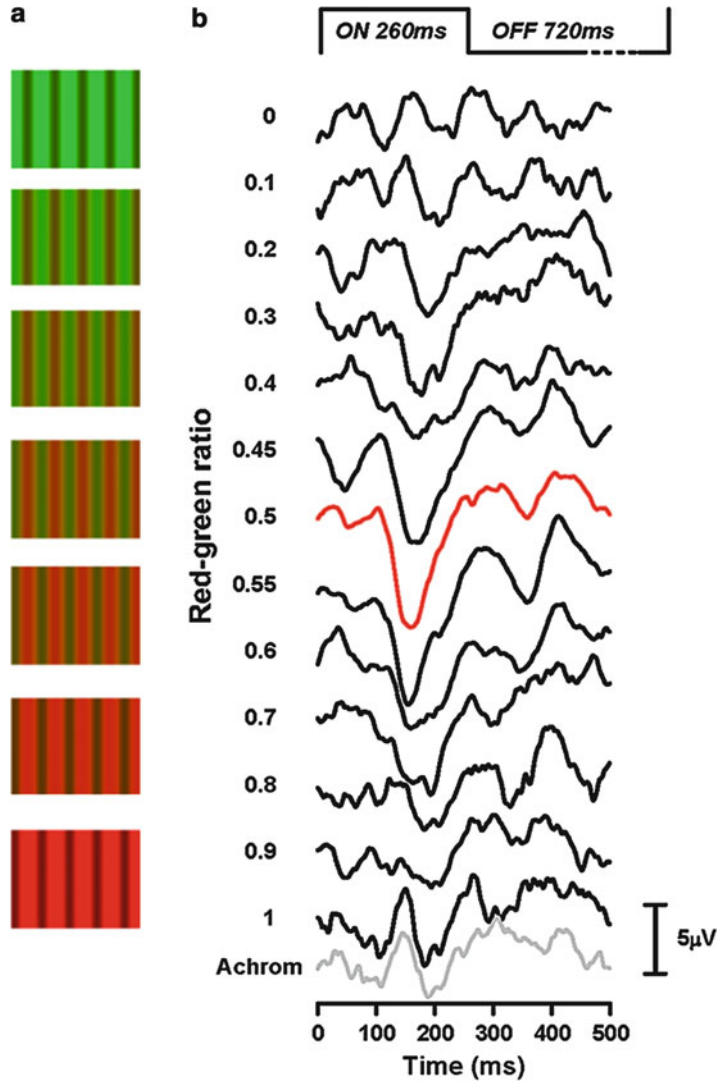


Visual Evoked Potentials, Fig. 3 Combining red and green gratings in phase to make an achromatic (yellow) stimulus and out of phase to make a chromatic (red-green) grating. Also shown is the offset phase when contrast is zero; note that this is the same for both stimuli

nonvisual or “reference” site (e.g., the vertex or an earlobe). In fact, the terms “reference” and “active” are misnomers as it is possible to record some level of visually related activity from anywhere on the scalp. An alternative name for this

Visual Evoked Potentials, Fig. 4

VEPs to a range of red-green ratios. The isoluminant response (when, for this subject, $RGR = 0.5$) is shown in red. The gray plot is the equivalent achromatic VEP, recorded to a yellow/black grating. Note that there is no significant difference between this response and those to the red/black or green/black gratings



montage is unipolar. A bipolar electrode montage places both electrodes over the same cortical area, a few cm apart. This generates smaller signals but has the advantage that distant sources of electrical “noise” tend to be of the same magnitude at both electrodes and are therefore not as readily recorded using a differential amplifier.

Chromatic Evoked Potentials

A spatial stimulus is defined in terms of many parameters, including its size (how many degrees of visual angle it subtends on the retina), the pattern type (e.g., sinusoidal or square-wave

gratings), its periodicity or spatial frequency (i.e., number of spatial cycles per degree of visual angle), the type of temporal modulation (e.g., sinusoidal or square-wave) and temporal frequency (number of temporal cycles per second). In order to elicit a chromatic response, the stimulus has to vary in color alone, in both spatial and temporal domains. It is important to achieve activation of color opponent mechanisms (see Fig. 2).

The simplest and most well-established method is to use a sinusoidal grating composed of two colors, say red and green (Fig. 3), which has been adjusted for each subject so that it is isoluminant. This means that, for an individual,

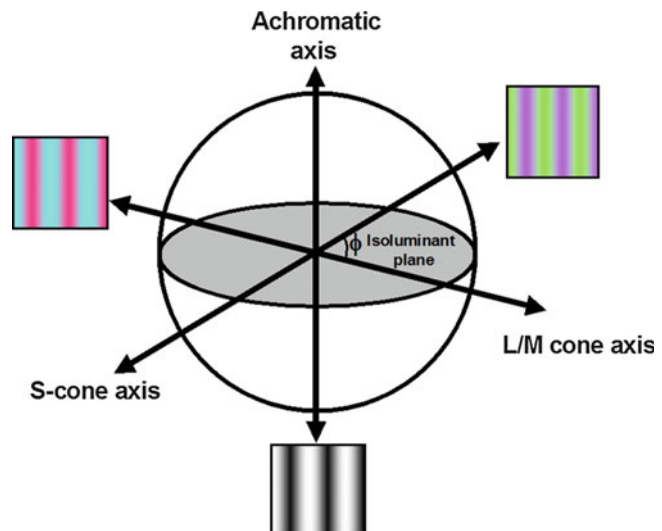
there is no change in luminance across the stimulus, but only a change in color or chromaticity. This is usually achieved with a psychophysical procedure such as heterochromatic flicker photometry (HFP), in which the two colors are exchanged at a fairly high rate (about 16Hz), and their luminance reciprocally adjusted until the perception of flicker (which results from residual luminance contrast) is minimised. In on-off mode, the pattern would be replaced with a plain yellow field of the same space-averaged hue and luminance.

In Fig. 4, a series of VEPs are shown in which luminance contamination is present in varying quantities. At isoluminance (the middle trace, shown in red), the onset response takes on a different shape, the achromatic positivity (seen at top and bottom of the figure) being replaced by a chromatic-specific negativity [1–5]. This achromatic-chromatic difference is not seen when checkerboards or reversal presentation are employed, and this is possibly because these stimuli are not optimal for the chromatic visual system, which, at least for red-green stimuli, is handled by the parvocellular (P) visual pathway [6]. The P pathway is dominated by sustained nerve fibers, whilst reversal favors transient responses, and perhaps therefore activates preferentially the magnocellular system [7].

Although red-green stimuli are relatively easy to generate on a graphics display using the red and green primary phosphors, they usually contain significant contrast in both red-green and blue-yellow color opponent pathways. Since these properties are handled by different retino-cortical pathways (respectively the parvocellular and koniocellular streams), it is desirable to optimize the stimulus so that it activates a particular pathway. Physical colors are often represented in a 3-dimensional space, an example of which is given in Fig. 5. See [► [Psychological Color Space and Color Terms](#)] for more on color spaces. The two component colors of the stimulus are selected from this space, forming a vector. This vector can be located at a particular orientation in order to restrict activation to one channel (say blue-yellow), whilst keeping the others (red-green and luminance) constant. In Fig. 6, a range of VEPs have been recorded to onset of chromatic stimuli which all lie on the isoluminant plane in MBDKL space [8, 9]; thus the luminance channel is held constant. The vector has been rotated for each successive recording so that at some point it passes through each color channel. When the chromatic axis (ϕ) is 0–180°, the stimulus is modulated along the L/M cone axis (“red-green”); when $\phi = 90$ –270°, it preferentially activates the S-cone axis (“yellow-blue”). 10 additional recordings have been made at intermediate

Visual Evoked Potentials,

Fig. 5 MBDKL color space; with thanks to Prof D. McKeefry

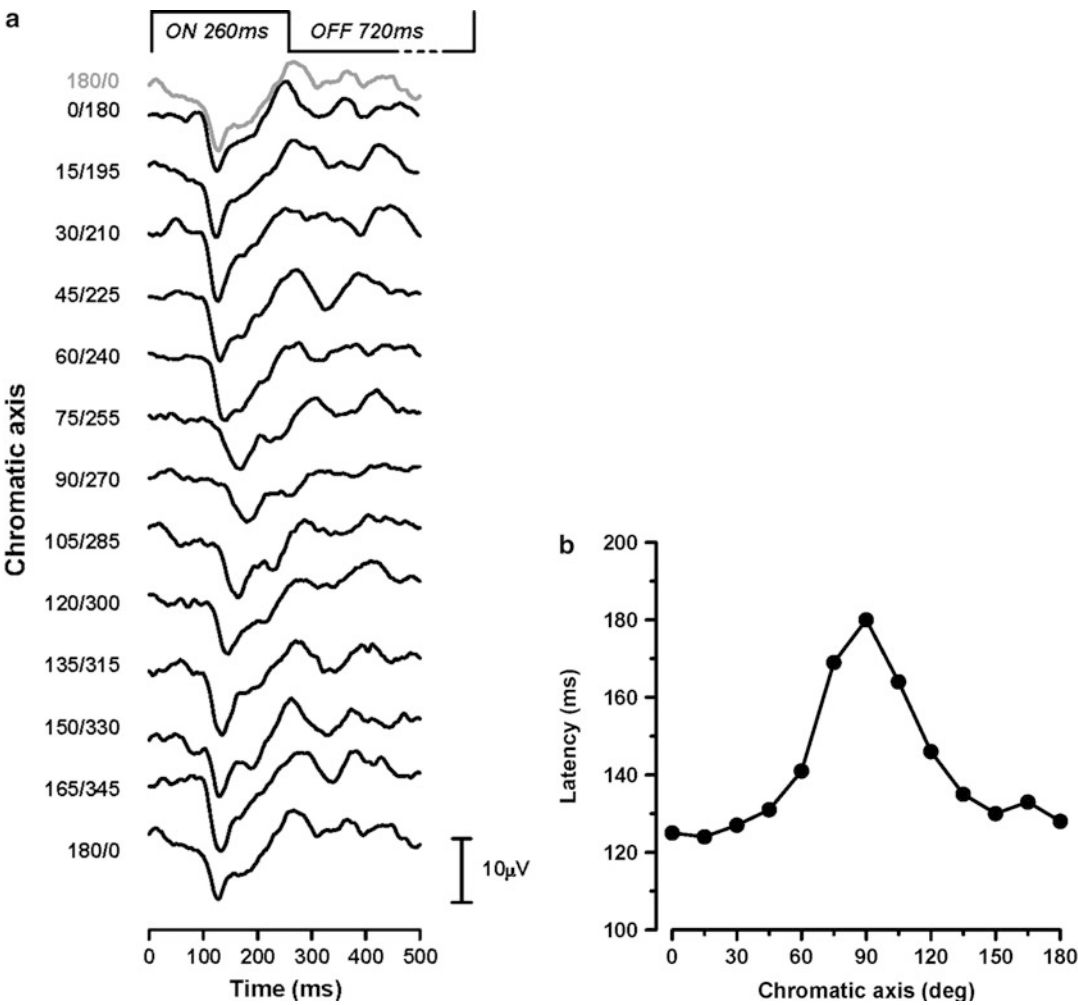


angles where one would expect a mixture of red-green and blue-yellow signals. The blue-yellow VEP, which taps the sluggish koniocellular pathway, has a much longer latency (time from onset of the stimulus to peak of the component) than the M or L-cone mediated responses (see Fig. 6 inset).

Limitations

There are a few factors which need to be taken into account when recording chromatic evoked potentials. Firstly, these stimuli may be contaminated

by luminance information due to chromatic aberrations. The key to avoiding the many complications arising from chromatic aberrations is to use low spatial frequency (<5 c/deg) sinusoidal gratings and to restrict the number of cycles visible. There is another good reason to use low spatial frequencies: the chromatic visual system is tuned to these, showing a low-pass tuning curve with a half-height of about 3 c/deg and an upper limit of about 10 c/deg [10]. The issue of stimulus size becomes particularly important when blue-yellow stimuli are used. This is because the central retina contains a higher concentration of yellow



Visual Evoked Potentials, Fig. 6 (Left) Occipital VEPs recorded to a range of chromatic axes, where 0/180 is cardinal red-green and 90/270 is cardinal blue-yellow

(see Fig. 5). (Below) Latency of the chromatic negativity as a function of chromatic axis

macular pigment, meaning that it is not possible to set a single value of isoluminance across the whole stimulus [11, 12]. Isoluminance also varies from subject to subject, even in those with normal color vision, largely because of the high inter-subject variability in the proportion of long-wavelength to medium-wavelength photoreceptors [► [Cone Fundamentals](#)]. Thus it is always advisable to establish isoluminance for each subject tested. The presence of color vision deficiencies is also important of course. In extremis, dichromatic individuals lack red-green opponency [Color Vision, Opponent Theory], and therefore do not generate red-green chromatic evoked potentials (see [Uses](#), below).

Uses

What benefits do chromatic VEPs bring, aside from giving us a tool to help understand the visual system? VEPs are useful in assessing the vision of preverbal individuals, and several studies have used VEPs to study the development of color vision in infants [13, 14]. There are many potential clinical uses, from the diagnosis of color vision deficiencies to the assessment of drug toxicity [15–18]. Studies have applied the technique to measure the effect of environmental toxins, which might be expected to preferentially affect the highly sensitive color visual system [► [Color Perception and Environmentally Based Impairments](#)]. The chromatic VEP has proven to be of value in neuro-ophthalmology [19–24], again because diseases such as optic neuritis seem to produce defective color vision. This may not be because the disease specifically targets color vision (for instance via the parvocellular pathway), but because the chromatic system is simply more sensitive to insult or injury, showing their effects more readily than the more robust achromatic visual system.

Cross-References

► [Color Perception and Environmentally Based Impairments](#)

► [Cone Fundamentals](#)

► [Psychological Color Space and Color Terms](#)

References

1. Carden, D., Kulikowski, J.J., Murray, I.J., Parry, N.R. A.: Human occipital potentials evoked by the onset of equiluminant chromatic gratings. *J. Physiol.* **369**, 44P (1985)
2. Murray, I.J., Parry, N.R.A., Carden, D., Kulikowski, J.J.: Human visual evoked potentials to chromatic and achromatic gratings. *Clin. Vis. Sci.* **1**, 231–244 (1987)
3. Berninger, T.A., Arden, G.B., Hogg, C.R., Frumkes, T.E.: Separable evoked retinal and cortical potentials from each major visual pathway: preliminary results. *Br. J. Ophthalmol.* **73**, 502–511 (1989)
4. Rabin, J., Switkes, E., Crognale, M., Schneek, M.E., Adams, A.J.: Visual-evoked potentials in 3-dimensional color space – correlates of spatiochromatic processing. *Vis. Res.* **34**, 2657–2671 (1994). [https://doi.org/10.1016/0042-6989\(94\)90222-4](https://doi.org/10.1016/0042-6989(94)90222-4)
5. Porciatti, V., Sartucci, F.: Normative data for onset VEPs to red-green and blue-yellow chromatic contrast. *Clin. Neurophysiol.* **110**, 772–781 (1999)
6. Lee, B.B.: Visual pathways and psychophysical channels in the primate. *J. Physiol.* **589**, 41–47 (2011). <https://doi.org/10.1113/jphysiol.2010.192658>
7. Kulikowski, J.J., Parry, N.R.A.: Human occipital potentials evoked by achromatic or chromatic checkerboards and gratings. *J. Physiol.* **388**, 45P (1987)
8. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
9. MacLeod, D.I.A., Boynton, R.M.: Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am.* **69**, 1183 (1979)
10. Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol.* **359**, 381–400 (1985)
11. Robson, A.G., Parry, N.R.A.: Measurement of macular pigment optical density and distribution using the steady-state visual evoked potential. *Vis. Neurosci.* **25**, 575–583 (2008)
12. Parry, N.R.A., Robson, A.G.: Optimization of large field tritan stimuli using concentric isoluminant annuli. *J. Vis.* **12**(12), 11 (2012). <https://doi.org/10.1167/12.12.11>
13. Parry, N.R.A., Murray, I.J.: Electrophysiological investigation of adult and infant colour vision deficiencies. In: Dickinson, C.M., Murray, I.J., Carden, D. (eds.) *John Dalton's Colour Vision Legacy*. Taylor Francis (1997)
14. Crognale, M.A.: Development, maturation, and aging of chromatic visual pathways: VEP results. *J. Vis.* **2**(6), 2 (2002). <https://doi.org/10.1167/2.6.2>

15. Kelly, J.P., Crognale, M.A., Weiss, A.H.: ERGs, cone-isolating VEPs and analytical techniques in children with cone dysfunction syndromes. *Doc. Ophthalmol.* **106**, 289–304 (2003). <https://doi.org/10.1023/a:1022909328103>
16. Tobimatsu, S., Celesia, G.G.: Studies of human visual pathophysiology with visual evoked potentials. *Clin. Neurophysiol.* **117**, 1414–1433 (2006). <https://doi.org/10.1016/j.clinph.2006.01.004>
17. Gomes, B.D., et al.: Normal and dichromatic color discrimination measured with transient visual evoked potential. *Vis. Neurosci.* **23**, 617–627 (2006). <https://doi.org/10.1017/s0952523806233194>
18. Pompe, M.T., Kranjc, B.S., Breceelj, J.: Chromatic VEP in children with congenital colour vision deficiency. *Ophthalmic Physiol. Opt.* **30**, 693–698 (2010). <https://doi.org/10.1111/j.1475-1313.2010.00739.x>
19. Russell, M.H.A., Murray, I.J., Metcalfe, R.A., Kulikowski, J.J.: The visual deficit in multiple sclerosis: a combined psychophysical and electrophysiological investigation. *Brain.* **114**, 2419–2435 (1991)
20. Porciatti, V., Sartucci, F.: Retinal and cortical evoked responses to chromatic contrast stimuli – specific losses in both eyes of patients with multiple sclerosis and unilateral optic neuritis. *Brain.* **119**, 723–740 (1996). <https://doi.org/10.1093/brain/119.3.723>
21. Tobimatsu, S., Kato, M.: Multimodality visual evoked potentials in evaluating visual dysfunction in optic neuritis. *Neurology.* **50**, 715–718 (1998)
22. Horn, F.K., Bergua, A., Junemann, A., Korth, M.: Visual evoked potentials under luminance contrast and color contrast stimulation in glaucoma diagnosis. *J. Glaucoma.* **9**, 428–437 (2000)
23. Sartucci, F., Murri, L., Orsini, C., Porciatti, V.: Equiluminant red-green and blue-yellow VEPs in multiple sclerosis. *J. Clin. Neurophysiol.* **18**, 583–591 (2001). <https://doi.org/10.1097/00004691-200111000-00010>
24. Wu, F., Yang, Y., Li, H., Odom, J.V.: Relationship of chromatic visual-evoked potentials and the changes of foveal photoreceptor layer in central serous chorioretinopathy patients. *Ophthalmic Physiol. Opt.* **31**, 381–388 (2011). <https://doi.org/10.1111/j.1475-1313.2011.00839.x>

Visual Illusions

- [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Visual Pigments

- [Photoreceptors, Color Vision](#)

Visual Pollution

- [Color Pollution](#)

VLC

- [Visible Light Communication](#)

von Bezold Spreading Effect

- [Assimilation](#)

Vora Measure, ν

- [Vora Value](#)

Vora Value

Poorvi L. Vora¹ and H. Joel Trussell²

¹Department of Computer Science, The George Washington University, Washington, DC, USA

²Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

Synonyms

[Vora measure, \$\nu\$](#)

Definition

The *Vora Value*, ν , measures the recording accuracy of a set of color filters [1]. The definition of ν takes into account the viewing and recording illuminants and the target color space. The original definition is independent of the data set to be recorded, does not take into account recording

noise, and is based on the Euclidean distance measure in the target color space.

Let $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s]$ be an $N \times s$ matrix whose columns are the vectors (or functions) that define the target color space for a specified viewing illuminant. If we assume that the vectors of $\mathbf{V}$ are independent, then they define a basis for the span of $\mathbf{V}$, $\text{Span}(\{\mathbf{v}_i\}_{i=1}^s)$. To simplify notation, we denote the span of $\mathbf{V}$ by $\mathbf{R}_s(\mathbf{V})$, where the subscript denotes the dimension of the vector space. For example, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ could represent the three *CIE XYZ* matching functions for viewing illuminant D65. Let the column vector $\mathbf{f}$ of length N represent the reflectance spectrum of the sample to be measured. Then the target color space is $\mathbf{R}_s(\mathbf{V})$, and the desired s -stimulus vector is

$$\mathbf{t} = \mathbf{V}^T \mathbf{f}$$

In our example, $\mathbf{t}$ would represent the *CIE XYZ* values of $\mathbf{f}$ for viewing illuminant D-65. The value of s need not be restricted to $s = 3$. For example, one may use $s = 6$ to measure the *CIE XYZ* values of $\mathbf{f}$ for viewing illuminants D-50 and D-65. For another example, $s = 4$ may be used to aid in color correction. For hyperspectral cameras and other multiband image recording systems, s may take on other, larger, values.

Let $\mathbf{M} = [\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_r]$ be an $N \times r$ matrix whose columns represent the effective recording filters (combination of the scanning filters and scanning illuminant). If recording noise can be ignored, the vector of recorded values is

$$\mathbf{g} = \mathbf{M}^T \mathbf{f}$$

It is typically physically impossible – or, at least, impractical – to construct vectors $\mathbf{m}_i = \mathbf{v}_i$ that match the ideal color matching functions exactly. Further, recording and viewing illuminants generally differ. In general, $r \neq s$, $\mathbf{M} \neq \mathbf{V}$, and $\mathbf{t} \neq \mathbf{g}$. The vector $\mathbf{g}$ is typically color corrected to obtain the best linear estimate of $\mathbf{t}$, $\mathbf{t}_{\text{est}} = \mathbf{A}\mathbf{g}$, where color correction matrix $\mathbf{A}$ is chosen so as to minimize the mean square estimation error, $E\{\|\mathbf{t} - \mathbf{t}_{\text{est}}\|^2\}$, where $E\{\cdot\}$ is the expected value operator. If the components of $\mathbf{f}$ are assumed to be

independent, identically distributed with variance σ^2 , it can be shown that

$$\mathbf{A} = \mathbf{V}^T \mathbf{M} (\mathbf{M}^T \mathbf{M})^{-1}$$

and

$$E\{\|\mathbf{t} - \mathbf{t}_{\text{est}}\|^2\} = \sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V} - \mathbf{V}^T \mathbf{M} (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{V})$$

The maximum value of $E\{\|\mathbf{t} - \mathbf{t}_{\text{est}}\|^2\}$ is $\sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V})$.

The Vora Value, ν , is a measure of the effectiveness of the set of recording filters $\{\mathbf{m}_i\}_{i=1}^r$, in recording the color of a sample in the target color space, $\mathbf{R}_s(\mathbf{V})$. It measures how accurately $\mathbf{t}$ can be estimated from $\mathbf{g}$ after linear color correction and is defined as

$$\nu = \left(\sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V}) - E\{\|\mathbf{t} - \mathbf{t}_{\text{est}}\|^2\} \right) / \sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V})$$

Note that ν increases linearly with a decrease in mean square estimation error and ranges in value from zero to one, where the maximum value of one corresponds to perfect recording.

The definition of ν does not take into account recording noise, real data sets, or perceptual error measures, each of which would influence the error expression. For measures that take these into account, please see the section on *Other Measures* below. The original form of ν is simple and provides an effective rough estimate of the accuracy of the filter set.

Motivation: The Measure ν and the Q-Factor

To understand the properties of a good filter set, note that $\mathbf{t}$ consists of the inner products (dot products) of the vectors $\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s$ with the vector $\mathbf{f}$ [2]. Thus $\mathbf{t}$ depends only on the *fundamental* of $\mathbf{f}$, which is defined as $\mathbf{P}_V(\mathbf{f})$, its orthogonal projection onto the target color space. On the other hand, any components of $\mathbf{f}$ that are orthogonal to all of

$\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s$ will not contribute to $\mathbf{t}$. Thus, in particular, $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ does not contribute to $\mathbf{t}$.

$$\mathbf{t} = \mathbf{V}^T \mathbf{f} = \mathbf{V}^T \mathbf{P}_V(\mathbf{f}) + \mathbf{V}^T (\mathbf{f} - \mathbf{P}_V(\mathbf{f})) = \mathbf{V}^T \mathbf{P}_V(\mathbf{f}) + \mathbf{0}$$

Figure 1a illustrates this for $s = 2$. Vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ define the target color space, the horizontal plane. The reflectance spectrum is represented by three-dimensional vector $\mathbf{f}$. The vector $\mathbf{t}$ is completely determined by $\mathbf{P}_V(\mathbf{f})$, the orthogonal projection of $\mathbf{f}$ onto the horizontal plane. The inner product of $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ with either of $\mathbf{v}_1$ or $\mathbf{v}_2$ is zero and hence does not contribute to $\mathbf{t}$.

Let $\mathbf{m}$ represent a single effective recording filter, obtained by combining the filter transmission function and the recording illuminant. Ignoring recording noise, the value $\mathbf{m}^T \mathbf{f}$ or the inner product of vectors $\mathbf{m}$ and $\mathbf{f}$ is the recorded value. The component of $\mathbf{m}$ that is orthogonal to the target color space, $\mathbf{m} - \mathbf{P}_V(\mathbf{m})$, contributes to inaccuracy in the recording of $\mathbf{t}$ because it measures energy in $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$. This is illustrated in Fig. 1b.

$$\begin{aligned} g = \mathbf{m}^T \mathbf{f} &= \left(\mathbf{P}_V(\mathbf{m})^T + (\mathbf{m} - \mathbf{P}_V(\mathbf{m}))^T \right) \\ &\quad (\mathbf{P}_V(\mathbf{f}) + (\mathbf{f} - \mathbf{P}_V(\mathbf{f}))) \\ &= \mathbf{P}_V(\mathbf{m})^T \mathbf{P}_V(\mathbf{f}) \\ &\quad + (\mathbf{m} - \mathbf{P}_V(\mathbf{m}))^T (\mathbf{f} - \mathbf{P}_V(\mathbf{f})) \end{aligned}$$

Thus we have seen that only the fundamental of $\mathbf{m}$, $\mathbf{P}_V(\mathbf{m})$ is useful for recording accuracy. The

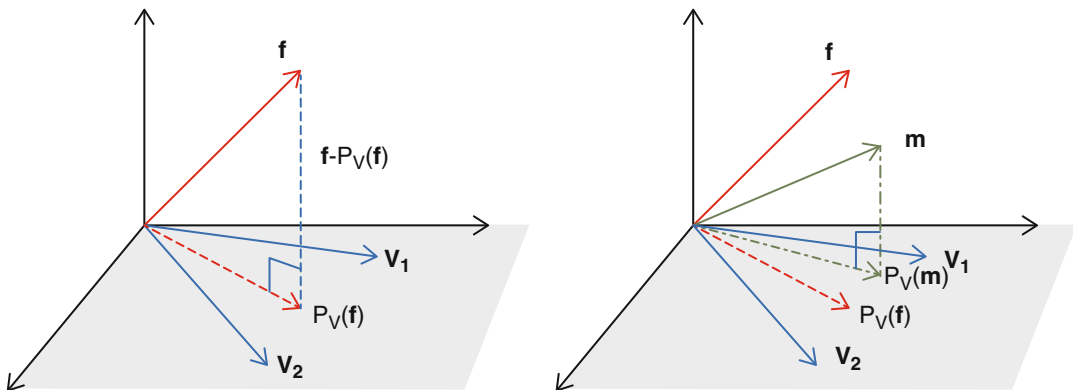
Colorimetric Quality Factor (CQF or q-factor, defined by Neugebauer in 1956) of $\mathbf{m}$ measures the accuracy of $\mathbf{m}$ from this perspective, and is defined as the energy in the vector $\mathbf{P}_V(\mathbf{m})$, as a fraction of the total energy in $\mathbf{m}$.

$$q(\mathbf{m}) = \|\mathbf{P}_V(\mathbf{m})\|^2 / \|\mathbf{m}\|^2$$

The value of $q(\mathbf{m})$ is larger when $\mathbf{P}_V(\mathbf{m})$ is a larger fraction of $\mathbf{m}$. If $\mathbf{m} \in \mathbf{R}_s(\mathbf{V})$, $\mathbf{P}_V(\mathbf{m}) = \mathbf{m}$, and $\mathbf{m}$ has a maximum q-factor of one.

While the CQF captures an important requirement of color scanning filters, recording filters are not used by themselves and it is necessary to evaluate filter sets as opposed to single filters. As in the case of a single filter, it is important for the filter set to measure the energy in $\mathbf{P}_V(\mathbf{f})$. Thus the approach underlying the q-factor is useful. However, while a single filter can record energy only in one “direction” of the target color space (e.g., a single filter can measure one of the CIE tristimulus values) the purpose of a filter set is to measure energy in all the directions of the target color space. Thus, for example, if we wish to record CIEXYZ tristimulus values, the filter set should accurately measure X, Y, and Z values. An evaluation criterion for a set of recording filters should hence evaluate whether it measures all the energy in $\mathbf{P}_V(\mathbf{f})$ — that is, whether it measures the energy in $\mathbf{P}_V(\mathbf{f})$ in all directions of the target color space.

The generalization of the q-factor to the evaluation of a filter set is not trivial. To see this,



Vora Value, Fig. 1 (a) $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ does not contribute to $\mathbf{t}$. (b) $\mathbf{m} - \mathbf{P}_V(\mathbf{m})$ measures parts of $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$

consider, for example, a set of three perfect red filters, each with a q-factor of one. While these are clearly inadequate to measure tristimulus values, *any* function of individual q-factors will indicate that the set is perfect. This limitation is illustrated in Fig. 2a for $s = 2$. The two filters $\mathbf{m}_1$ and $\mathbf{m}_2$ are both perfect, lie in $\mathbf{R}_r(\mathbf{M})$ and have q-factors of 1. However, one is a multiple of the other. Hence the filter set is not capable of measuring that part of $\mathbf{P}_V(\mathbf{f})$ that is orthogonal to the filters. It is hence not capable of measuring all the energy in $\mathbf{P}_V(\mathbf{f})$.

The general case is illustrated for $s = r = 2$ in Fig. 2a. The two recording filters $\mathbf{m}_1$ and $\mathbf{m}_2$ do not lie in the target color space, which is the horizontal plane, but in the shaded plane. The desired recorded values depend on $\mathbf{P}_V(\mathbf{f})$. The values that are recorded in practice depend on $\mathbf{P}_M(\mathbf{f})$.

$$\begin{aligned} \mathbf{g} &= \mathbf{M}^T \mathbf{f} = \mathbf{M}^T \mathbf{P}_M(\mathbf{f}) + \mathbf{M}^T (\mathbf{f} - \mathbf{P}_M(\mathbf{f})) \\ &= \mathbf{M}^T \mathbf{P}_M(\mathbf{f}) + \mathbf{0} \end{aligned}$$

Hence the set of filters is perfect when $\mathbf{R}_r(\mathbf{M}) = \mathbf{R}_s(\mathbf{V})$, and the error is not dependent on individual filters but on the spaces. Thus $E\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\}$ depends on how much of the energy of $\mathbf{R}_r(\mathbf{M})$ is contained in $\mathbf{R}_s(\mathbf{V})$, or, equivalently, how much of the energy of orthogonal filters spanning $\mathbf{R}_r(\mathbf{M})$ is contained in $\mathbf{R}_r(\mathbf{V})$. It may be shown that

$$v(\mathbf{V}, \mathbf{M}) = \sum_{i=1}^r \frac{q(\mathbf{o}_i)}{s},$$

where $\{\mathbf{o}_i\}_{i=1}^r$ is an orthogonal basis for $\mathbf{R}_r(\mathbf{M})$ and $q(\mathbf{o}_i)$ is the q-factor of $\mathbf{o}_i$. The use of orthogonal filters prevents correlation among filters from artificially increasing the value of the measure (as in Fig. 2a).

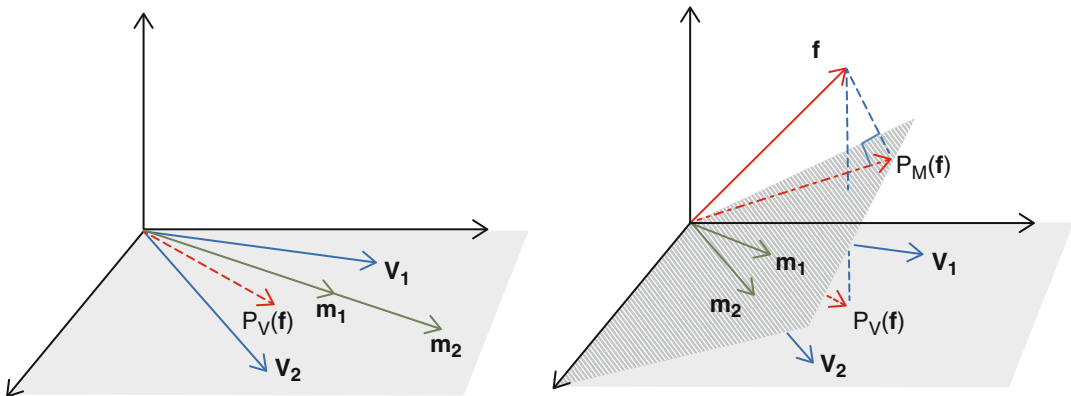
Additional insight into the power of the Vora Value is obtained by considering the case where the target space is the horizontal plane in 3-space. A set of three independent vectors, none of which lie in the plane, form a basis of the 3-space. None of the vectors will have a q-factor of one. It is clear that the span of the three vectors includes the target space. For this case, $v = 1$, as it should. Thus, we can create a filter set that produces perfect tristimulus values from imperfect filters. The cost of this is that it requires more filters than the dimension of the target color space, i.e., $r > s$.

The Measure v and Principal Angles

It may be shown that

$$v(\mathbf{V}, \mathbf{M}) = \sum_{i=1}^s \frac{\cos^2(\theta_i)}{s}$$

where θ_i is the i^{th} principal angle [3] between the target color space, $\mathbf{R}_s(\mathbf{V})$, and $\mathbf{R}_r(\mathbf{M})$.



Vora Value, Fig. 2 (a) *Left*: Two or more scanning filters should not record energy in a single direction. (b) *Right*: Differences between $\mathbf{R}_r(\mathbf{M})$ and $\mathbf{R}_s(\mathbf{V})$ lead to inaccurate recording

Other Measures

The measure v is based on a mean-square error in the target color space assuming independent, identically distributed data. It does not take into account real data sets, perceptual error, or recording noise.

Finlayson and Drew [4, 5] address the issue of real data sets. They propose a similar measure based on the *White Point Preserving Vora Error With Maximum Ignorance With Positivity (WPPVEMIP)*, which assumes that one is maximally ignorant of the distribution of $\mathbf{f}$ except for knowing that all values in the vector $\mathbf{f}$ are positive and that the color correction procedure preserves white points. They demonstrate that the *WPPVEMIP* is a more accurate estimate of the real error, and hence that the corresponding measure is more accurate than v .

Wolski et al. [6] address the issue of perceptual error by using a linear approximation of CIE Lab error in the neighborhood of the points in the (known) data set. The filter sets designed by this method minimize an error measure that is a better estimate of perceptual error than the mean square error used to define v .

Sharma and Trussell [7] address the issue of recording noise and propose a *Figure of Merit (FOM)* based on the approach of Wolski et al.

The FOM is derived from an error that takes into consideration recording noise.

References

1. Vora, P.L., Joel Trussell, H.: Measure of goodness of a set of color scanning filters. *J. Opt. Soc. Am. A.* **10**(7), 1499–1508 (1993)
2. Neugebauer, H.E.J.: Quality factor for filters whose spectral transmittances are different from color mixture curves, and its application to color photography. *J. Opt. Soc. Am.* **46**(10), 821–884 (1956)
3. Golub, G.H., Van Loan, C.F.: *Matrix Computations*. Johns Hopkins University Press, Baltimore (1996)
4. Finlayson, G.D., Drew, M.S.: The maximum ignorance assumption with positivity. In: *4th Color Imaging Conference: Color, Science, Systems And Applications, IS&T/SID*, pp. 202–205 (1996)
5. Finlayson, G.D., Drew, M.S.: White-point preservation enforces positivity. In: *6th Color Imaging Conference: Color, Science, Systems And Applications*, pp. 47–52 (1998)
6. Mark, W., Bouman, C.A., Allebach, J.P., Eric, W.: Optimization of sensor response functions for colorimetry of reflective and emissive objects. *IEEE Trans. Image Proc.* **5**(3), 507–517 (1996)
7. Sharma, G., Trussell, H.J.: Figures of merit for color scanners. *IEEE Trans. Image Proc.* **6**(7), 990–1001 (1997)

VT and Color

► [Vantage Theory of Color](#)

Wax-Resist Dyeing

► [Batik](#)

Wet or Chemical Finishes

► [Finish, Textile](#)

Wide-Dynamic-Range Imaging

► [High Dynamic Range Imaging](#)

Wood's Lamps

► [Blacklight Lamp](#)

World Color Survey

Paul Kay and Richard S. Cook
Department of Linguistics, University of
California, Berkeley, CA, USA

Definition

The World Color Survey collected comprehensive color-naming data from an average of 24 speakers

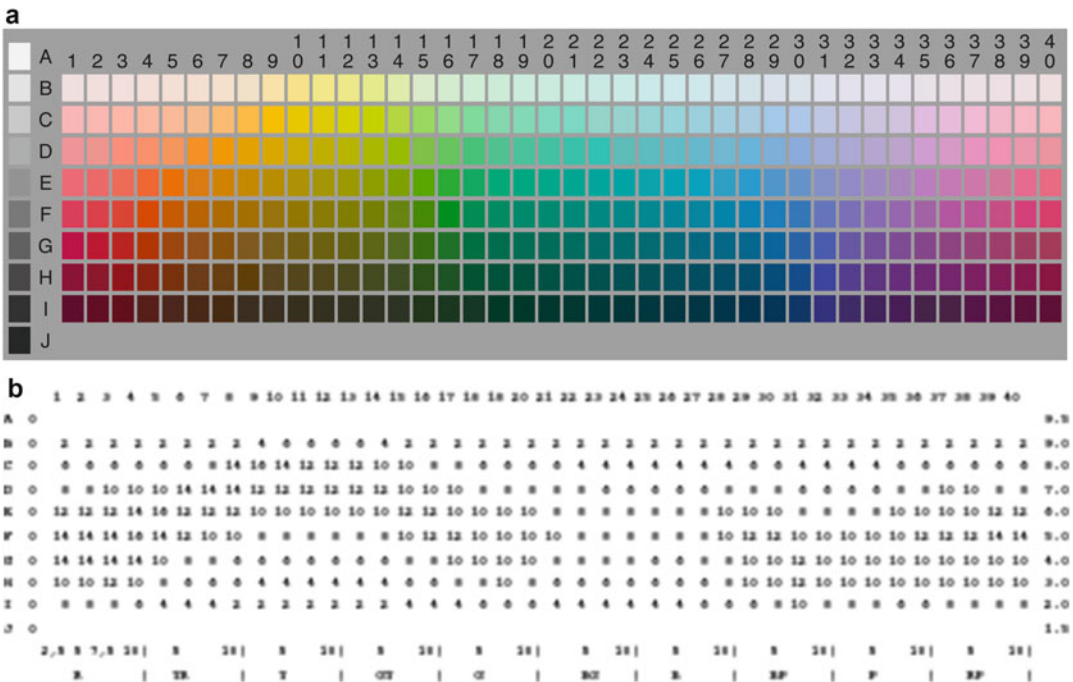
of each of 110 unwritten languages from around the world. Analysis of these data has resulted in a number of research publications. The data are available at <http://www1.icsi.berkeley.edu/wcs/data.html>.

Introduction

The World Color Survey (WCS) was undertaken to investigate the main findings of Berlin and Kay (B&K) [1]. These were (A) that there exist universal crosslinguistic constraints on color naming and (B) that basic color terminology systems tend to develop in a partially fixed order. To this end, the WCS collected color-naming data from speakers of 110 unwritten languages. The WCS data are available in the WCS Data Archive. This entry reviews the history of the WCS, including the creation of the online data archive, and describes some recent uses of the archive to investigate constraints on color naming across languages.

The WCS: History and Methodology

The WCS was begun in 1976 to evaluate the findings of B&K in a full-scale field study. B&K had investigated the color terminology systems of 20 languages in the following way. The stimulus palette used by Erik Lenneberg and John Roberts [2], consisting of 320 Munsell chips of 40 equally



World Color Survey, Fig. 1 (a) The WCS stimulus palette. (b) Munsell and WCS coordinates for stimulus palette of (a). The leftmost column and the top row give the WCS coordinates for lightness and hue, respectively. The rightmost column and the bottom two rows give the

Munsell coordinates for value and hue, respectively. Entries in the body of the table show the corresponding Munsell chroma numbers. (With regard to the A and J rows, there are no Munsell hues at the extremes of value (lightness): 9.5 (*white*) and 1.5 (*black*))

spaced hues and eight levels of lightness (value) at maximum saturation (chroma) for each (hue, value) pair, was supplemented by nine Munsell achromatic chips (black through gray to white) – an approximation of the resulting stimulus palette is shown in Fig. 1a and the corresponding Munsell coordinates in Fig. 1b.

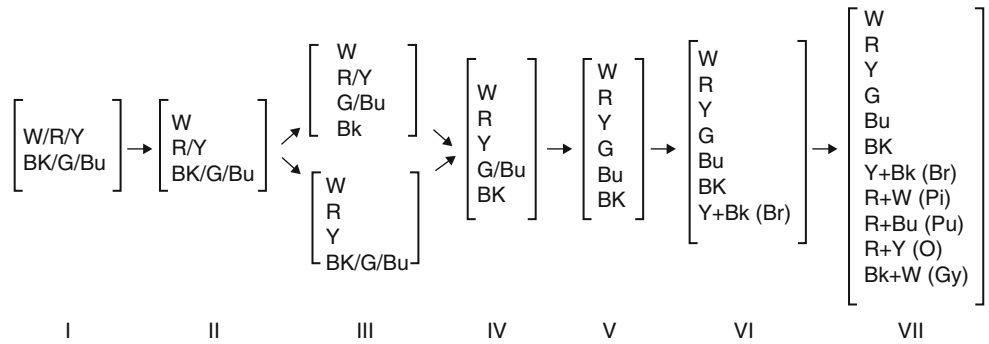
First, without the stimulus palette present, the major color terms of the collaborator’s native language were elicited by questioning that was designed to find the smallest number of simple words with which the speaker could name any color (*basic color terms*). Once this set of terms was established, the collaborator was asked to perform two tasks. In the *naming* task the stimulus palette was placed before the speaker, and for each color term *t*, a piece of clear acetate was placed over the stimulus board, and the collaborator was asked to indicate, with a grease pencil on the acetate sheet, all the chips that he or she could call *t*. In the *focus* task the stimulus palette was shown as before, and the collaborator was asked

to indicate the best example(s) of *t* for each basic color term *t*. B&K concluded that

[1] The referents for the basic color terms of all languages appear to be drawn from a set of eleven universal perceptual categories, and [2] these categories become encoded in the history of a given language in a partially fixed order. [1]

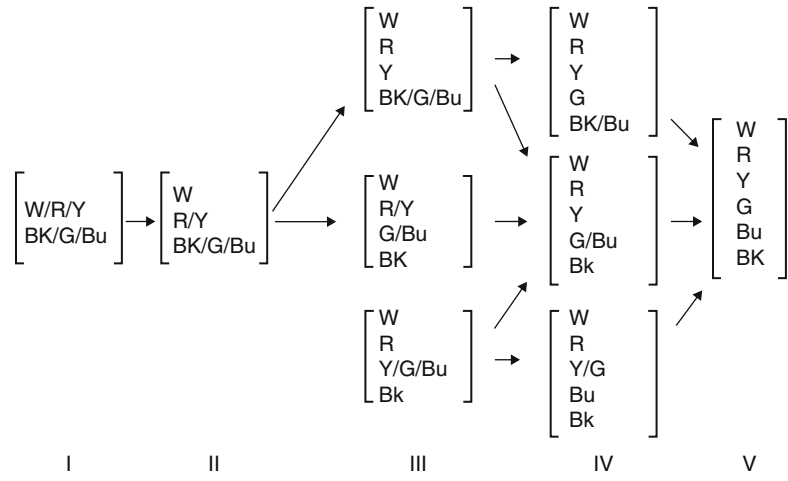
The original universal evolutionary sequence of color term development postulated by B&K is shown in Fig. 2. It has subsequently been revised in detail as more data has become available, but the main outlines of the original sequence have remained intact (see Fig. 3, for the most recent revision).

The B&K results were immediately challenged by anthropologists on the grounds that the sample of experimental languages was too small, too few collaborators per language were questioned (sometimes only one), all native collaborators also spoke English, the data were collected in the San Francisco Bay area rather than in the homelands of the target languages, many regions of the



World Color Survey, Fig. 2 Original B&K evolutionary sequence of color term development

World Color Survey,
Fig. 3 Revised
(2009) evolutionary
sequence of color term
development



world and language families were underrepresented or overrepresented in the sample of 20, and that the sample of 20 had too few unwritten languages of low technology cultures [3–6]. The results were nevertheless supported by various ethnographic and experimental studies conducted after 1969 and were from the start largely accepted by psychologists and vision researchers (e.g., [7–9]. See also [10], 498ff, [11], 133 ff).

Work on the WCS was begun in the late 1970s. Through the cooperation of SIL International (then the Summer Institute of Linguistics), which maintains a network of linguist-missionaries around the world, data on the basic color term systems of speakers of 110 unwritten languages representing 45 different families and several major linguistic stocks were gathered in

situ. Fieldworkers were provided with a kit containing the stimulus materials (330 individual chips in glass 35-mm slide sleeves for the naming task and the full stimulus palette for the focus task) as well as coding sheets on which to record collaborators’ responses. The included instructions requested that fieldworkers collect data from at least 25 speakers, both males and females, and urged them to seek out monolingual speakers insofar as possible. The modal number of speakers actually assessed per language was 25, and the average number was 24. (A facsimile of the WCS instructions to fieldworkers and of the original coding sheets is available on the WCS website.) The aim was to obtain names, category extent, and best examples of basic color terms in each language – basic color terms being described in the instructions as “the smallest set of simple

words with which the speaker can name any color.”

The WCS methodology coincided with that of the B&K study in the use of essentially the same set of Munsell color chips. One white chip was added in the WCS study that was whiter than any chip available at the time of the B&K study, making for a total of ten achromatic chips and an overall total of 330 chips, as shown in Fig. 1.

The WCS differed from B&K in the technique for eliciting naming responses. In the WCS procedure, no preliminary interview was administered to establish a set of basic color terms, and in the naming task the 330 individual color stimuli were shown to each cooperating speaker, one by one, according to a fixed, pseudorandom order, and a name elicited for each (in contrast with the B&K procedure of presenting the entire stimulus palette at once in eliciting naming responses). Fieldworkers were instructed to urge observers to respond with short names (although, depending on the morphology of the language, particular field circumstances, and local culture, there was considerable variation in the degree to which the field investigators were able to satisfy these desiderata). Identification of basic color terms, therefore, was done by the fieldworker as a result of the naming task itself, rather than through prior elicitation. The best example (focus) responses were elicited in the same way in both studies: once a set of basic color terms was isolated, the native observer was presented with the full palette and asked to indicate the chip or chips that represented the best example of each term, one by one.

Initial Analysis

Originally, the naming and best example data of the WCS were entered into separate files for each language; they were not compiled into a unified database until the early 2000s. In 2009, a monograph [13] appeared, based on analysis of data in this form. It contains a separate chapter for the color-naming system of each language, identifying the basic terms of the language through a variety of ways of summarizing and displaying the data. An updated version of the original B&K evolutionary sequence was postulated and is depicted in Fig. 3.

Uses of the WCS Archive

The WCS data archive has been used in investigating two broad questions, one concerning *universals* and other concerning *variation*, of color naming, corresponding to the two major conclusions of B&K [1]. Numerous statistical studies utilizing the WCS online have been conducted.

Universals of Color Naming

Since B&K found evidence for universals in color naming across languages, the existence of such constraints has generally been accepted in the scientific community. However, there have always been dissenters from this consensus (e.g., [2, 3]), and this dissenting view has recently gained prominence (e.g., [14–19]). Criticisms of the universalist position have come in two major varieties. The first points out that B&K’s findings were never objectively tested, as they relied on visual inspection of color-naming data. Lucy [15] challenges such a methodology as hopelessly subjective:

[Work in the B&K tradition] not only seeks universals, but sets up a procedure which guarantees both their discovery and their form. . . . when a category is identified . . . it is really the investigator who decides which ‘color’ it will count as . . . What appears to be objective - in this case, a statement of statistical odds - is [not]. ([15], p. 334)

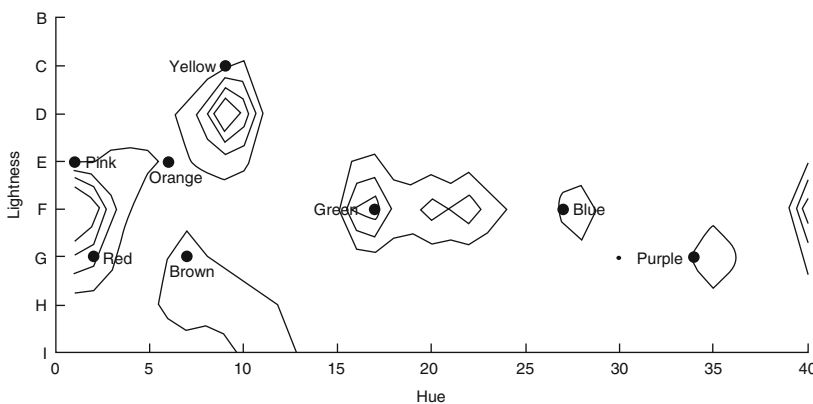
On this view, B&K’s subjective methodology allowed them to impose their own universalistic assumptions on their data – so the universals are actually in the minds of the investigators, not in the languages of the world. The second strand of criticism points out that B&K’s data were drawn primarily from written languages and thus may not be representative. This point is coupled with analyses of particular unwritten languages, which are claimed to counterexemplify universal constraints (e.g., Berinmo: [18, 19]; Hanunóo and Zuni: [15]). Subsequent, more detailed analyses of each of these languages have found that each fits the universal pattern [20]. Disputes of this sort over conflicting interpretations of individual color-naming systems could continue indefinitely.

Objective statistical studies were needed to resolve the issue.

The WCS database has been used in a number of independent statistical studies to test the hypothesis that there are statistical constraints on the basic color-naming systems of languages. The weight of the evidence supports the conclusion that such universal statistical constraints exist. The centroids of the naming responses to all color categories documented in the survey were found to cluster in color space more closely than chance would dictate; in a Monte Carlo simulation of the WCS, in which on each of 1,000 trials the modal color-naming pattern of every WCS language was rotated a random hue angle, the actual WCS naming centroids were found to cluster more tightly than the naming centroids in any of the 1,000 hypothetical (rotated) versions of the WCSs [21]. Figure 4 presents a contour plot of the WCS naming centroids compared to those of English.

The best example choices (foci) of all WCS color terms were found to also cluster more tightly than the centroids of naming categories, suggesting an intimate relation between “focal” colors and universal tendencies of color naming [23]; further, the WCS focal choices were found to cluster closely to those of English or other familiar written languages, as shown in Fig. 5.

Whereas the studies just discussed relied on color naming and focus choice data grouped by language, a clustering and concordance study based on the naming patterns of individual participants independently supports the conclusion that the WCS languages largely partition the color space in ways that, although often having fewer basic terms than English and hence fewer boundaries in their lexical “map” of color space, tend strongly to place boundaries in the same locations as do English and other familiar written languages [24]. This study also detected a hierarchical order in the lexical partitions of color space compatible with that depicted in Fig. 3. The question naturally arises regarding the degree to which WCS “focus” judgments, that is, participants’ judgments of the most typical examples of named categories, will agree with judgments of unique hues. For example, to what extent do people’s judgments of the best example of green agree with their (actually, in this case, with other people’s) judgments of a green that contains neither blue nor yellow. In a study that pruned the WCS data to consider 38 languages that yield unequivocal results for the Hering fundamental hues, red, yellow, green, and blue, it was found that the focal judgments of the WCS participant speakers of unwritten languages agreed well with unique hue judgments of 300 speakers of several written languages [25].

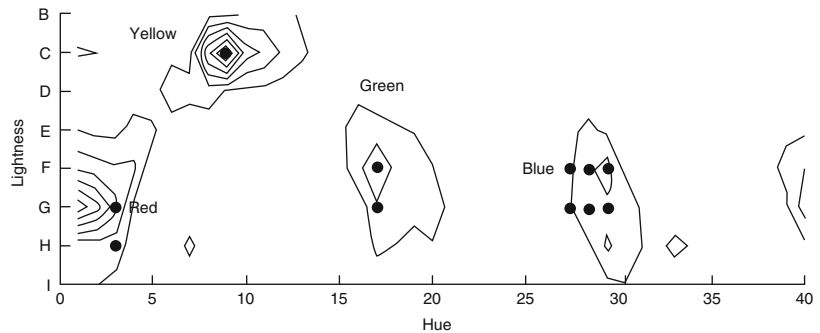


World Color Survey, Fig. 4 Contour plot of WCS speakers’ naming centroids, compared with English naming centroids (*black dots*); source for English naming centroids: [22]. The outermost contour represents a height of

100 centroids, and each subsequent contour represents an increment in height of 100 centroids (Source of figure: [21])

World Color Survey,

Fig. 5 Contour plot of focal color responses from WCS languages, with superimposed focal responses from English (B&K data, shown as dots), plotted against the stimulus palette (Source: [23])



Systematic, statistical comparison with the WCS database of languages claimed to violate universal color-naming tendencies has refuted that claim; however, a small number of other languages, not the object of such claims, have in fact been found to violate universal tendencies [26]. In a different study, languages with terms roughly equivalent to red, yellow, green, and blue nevertheless were found to differ slightly in their average focus placements, although this variation was greatly exceeded by that found among speakers of the same language [27]. A tentative explanation of the universal tendencies in color naming was found in a study that modeled hypothetical color-naming systems minimizing within-category distance in color space; good fit between the data generated by this model and the WCS data was achieved [28]. One explanation for variation across speakers within a given language is that there appear to be a small number of patterns of naming that occur among some speakers of many languages, with the speakers of few languages all following the same naming pattern [29]. Analysis of the WCS database has revealed that the categories named by basic color terms in the world's languages tend to be convex sets in color space [30]. An iterated learning study in which a stable color-naming system is achieved by interacting hypothetical agent has shown that equipping such hypothetical agents with the human just noticeable difference function at the start is sufficient to produce a final output that matches the WCS data well [31]. An iterated learning study using actual human learners showed that limiting a given simulation to just the number of color terms also produced systems

of color naming that were statistically close to corresponding systems in the WCS [32]. Extensive narrative descriptions of the color-naming systems of each WCS language, supported by charts and tables and keyed to the overall universal classification scheme, are available in monograph form [13].

Cross-References

- [Berlin and Kay Theory](#)
- [Bilingual Color Naming/Categories](#)
- [Color Categorical Perception](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effects of Color Terms on Color Perception and Cognition](#)
- [Infant Color Categories](#)
- [Unique Hues](#)
- [Vantage Theory of Color](#)

References

1. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley/Los Angeles (1969)
2. Lenneberg, E.H., Roberts, J.M.: *The Language of Experience: A study in Methodology*. Memoir 13 of International Journal of American Linguistics. (1956).
3. Durbin, M.: Review (1). *Semiotica*. **6**, 257–278 (1972)
4. Hickerson, N.: Review of (1). *Int. J. Am. Linguistics*. **37**, 257–270 (1971)
5. Collier, G.A.: Review of (1). *Language*. **49**, 245–248 (1973)
6. Conklin, H.C.: Color categorization: Review of (1). *Language*. **75**, 931–942 (1973) 1955 Hanunóo color categories. *Southwestern J. Anthropol.* **11**, 339–344

7. Brown, R.W.: Reference. *Cognition*. **4**, 125–153 (1976)
8. Miller, G.A., Johnson-Laird, P.: *Language and Perception*. Harvard University Press, Cambridge, MA (1976)
9. Ratliff, F.: On the psychophysiological bases of universal color terms. *Proc. Am. Philos. Soc.* **120**, 311–330 (1976)
10. Kaiser, P., Boynton, R.M.B.: *Human Color Vision*. Optical Society of America, Washington, DC (1996)
11. Boynton, R.M.: Insights gained from naming the OSA colors. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 135–150. Cambridge University Press, Cambridge (1997)
12. Collier, G.A., Dorfinger, G.K., Gulick, T.A., Johnson, D.L., McCorkle, C., Meyer, M.A., Wood, D.D., Yip, L.: Further evidence for universal color categories. *Language*. **52**, 884–890 (1976)
13. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
14. Lucy, J.A.: *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge University Press, Cambridge, UK (1992)
15. Lucy, J.A.: The scope of linguistic relativity: an analysis and review of empirical research. In: Gumperz, J.J., Levinson, S.C. (eds.) *Rethinking Linguistic Relativity*. Cambridge University Press, Cambridge, UK (1996)
16. Lucy, J.A.: The linguistics of color. In: Hardin, C.L., Luisa, M. (eds.) *Color Categories in Thought and Language*. Cambridge University Press, Cambridge, UK (1997)
17. Saunders, B.A.C., van Brakel, J.: Are there nontrivial constraints on colour categorization? *Brain Behav. Sci.* **20**, 167–228 (1997)
18. Davidoff, J., Davies, I.R.L., Roberson, D.: Colour categories in a stone-age tribe. *Nature*. **398**, 203–204 (1999)
19. Roberson, D., Davies, I.R.L., Davidoff, J.: Colour categories are not universal: replications and new evidence from a stone age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
20. Kay, P.: Methodological issues in cross-language color naming. In: Jourdan, C., Tuite, K. (eds.) *Language, Culture and Society*, pp. 115–134. Cambridge University Press, Cambridge (2006)
21. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
22. Sturges, J., Whitfield, T.W.A.: Locating basic colours in the Munsell space. *Color. Res. Appl.* **20**, 364–376 (1995)
23. Regier, T., Kay, P., Cook, R.: Focal colors are universal after all. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8386–8391 (2005)
24. Lindsey, D.T., Brown, A.M.: Universality of color names. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16609–16613 (2006)
25. Kuehni, R.G.: Nature and culture: an analysis of individual focal color choices in World Color Survey languages. *J. Cogn. Cult.* **7**, 151–172 (2007)
26. Kay, P., Regier, T.: Color naming universals: the case of Berinmo. *Cognition*. **102**, 289–298 (2007)
27. Webster, M., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R., Paramei, G., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
28. Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language*. **85**, 884–892 (2009)
29. Lindsey, D.T., Brown, A.M.: World Color Survey color naming reveals universal motifs and their within-language diversity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19785–19790 (2009)
30. Jäger, G.: Natural color categories are convex sets. In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (eds.) *Logic, Language and Meaning*. 17th Amsterdam Colloquium, pp. 11–20. Springer, Amsterdam (2009)
31. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2403–2407 (2010)
32. Xu, J., Griffiths, T. L., Dowman, M.: Replicating color term universals through human iterated learning. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Portland, Oregon: Cognitive Science Society (2010)

Wright, David

Robert W. G. Hunt

Department of Colour Science, University of
Leeds, Leeds, UK



W. David Wright was a British physicist and color scientist who made very important contributions to colorimetry and visual science. In fact he is generally regarded as one of the fathers of colorimetry as it is practiced today.

Wright graduated at Imperial College, London University, in 1926 and received a Ph.D. in 1929 and a D.Sc. in 1937. He was a Medical Research Council student at Imperial College from 1926 to 1929, and in this period, the spectral color-matching properties of ten observers were measured.

He was a Research Engineer at Westinghouse Electric and Manufacturing Co., Pittsburgh, USA, in 1929–30, where he undertook early research on color television.

On his return to England, at Imperial College, he became a Lecturer and Reader in Technical Optics in 1930–1951 and Professor of Applied Optics in 1951–1973. He was a Research and Consultant Physicist to Electrical and Musical Industries in 1930–1939. He became Kern Professor of Communications at the Rochester Institute of Technology in 1984–1985.

Main Interests

W. David Wright had a number of interests which included optics, vision, photometry, colorimetry, color perception, color applications, and color paintings. He supervised a succession of research students in color science, many of whom subsequently held senior positions in academia and industry.

Wright's most important research work was the measurement, for ten observers, of the way in which the colors of the spectrum are matched by beams of red, green, and blue light added together. This work, together with a similar study carried out by John Guild (at the National Physical Laboratory at Teddington, England, with seven additional observers), forms the basis of the international standard for measuring color established by the Commission Internationale de l'Éclairage (CIE). The quality of this experimental work was so high that the standard, although now more than eighty years old, is still in universal use.

Wright's researches also led, for the first time, to definitive descriptions of the main types of color deficiency (color blindness). His book *Researches on Normal and Defective Colour Vision*, published in 1946, summarizes the results of his research on color vision. He also wrote the book *The Measurement of Colour*, the four editions of which provided a widely used practical guide to colorimetry from 1944 to the present time. His leisure interests included an appreciation of paintings and religious service and church activities.

Wright's books and key publications include:

The Perception of Light, Blackie, 1938.
Researches on Normal and Defective Colour Vision, Kimpton, 1946.
The Measurement of Colour, Hilger, 1944 (2nd ed., 1958; 3rd ed., 1964; 4th ed., 1969).
Photometry and the Eye, Hatton, 1950.
The Rays are not Coloured, Hilger, 1967.

He also authored about 80 original papers, mainly dealing with color and vision.

Wright received numerous awards and was a member of several professional bodies which included:

The Physical Society Thomas Young Oration, 1951.
 Honorary D.Sc. City University, 1971.
 The AIC (International Colour Association) Judd Award, 1977.
 Honorary D.Sc. University of Waterloo, Canada, 1991.

He also provided public and professional service which included:

Founder of the Physical Society Colour Group, 1941.
 Chairman of the Physical Society Colour Group, 1941–1943.
 Vice-President of the Physical Society, 1948–1950.
 Secretary of the International Commission for Optics, 1953–1966.

Chairman of the Physical Society Optical Group, 1956–1959.
 President of the International Colour Association (AIC), 1967–1969.
 Chairman of the Colour Group (Great Britain), 1973–1975.

Wyszecki, Gunter

Rolf G. Kuehni
 Charlotte, NC, USA



Biography

Gunter Wyszecki was a German-Canadian mathematician/physicist who made important contributions to the fields of colorimetry, color discrimination, color order, and color vision [1]. He was born in Tilsit, East Prussia, Germany (today Sovetsk, Russia), in 1925. He attended the Technische Universität Berlin where he obtained a Dr.-Ing. degree, with a dissertation on normal and anomalous trichromacy [2]. In 1953 he was awarded a Fulbright Scholarship and for a year joined Deane B. Judd at the Colorimetry and Photometry Section of the US National Bureau of Standards in Washington, DC. In 1955 Wyszecki joined the National Research Council Canada in Ottawa where he became the leader of its Optics

Section in 1960 and assistant director of the Division of Physics in 1982 and where he remained until his untimely death from leukemia on June 22, 1985.

Major Accomplishments/Contributions

Wyszecki is best known for his scientific contributions to and leadership in the International Commission on Illumination (CIE). He was chairman of its Colorimetry Committee from 1963 to 1975, vice president of the organization from 1979 to 1983, and its president from 1983 until his death. During this period the CIE made many important recommendations in colorimetry, remaining valid today, such as 1 nm tables of the color-matching functions of the two CIE standard observers and the standard illuminants A and D65, addition of integrating sphere reflectance factor measurement as a recommended measuring geometry, the 1964 ($U^*V^*W^*$) and the 1976 CIELAB and CIELUV uniform color space and color-difference formulas, and others.

Metamerism Wyszecki introduced the important mathematical concept of “metameric blacks,” psychophysical definitions of blacks with tristimulus values 0, 0, 0 that, within limits, can be added to a spectral reflectance to form the various possible metamers having an identical set of tristimulus values under a given light [3]. With W.S. Stiles he also developed mathematical methods to calculate by various methodologies the number of possible metamers for given chromaticities, peaking at the achromatic colors [4].

Wyszecki seven-field colorimeter In 1965 Wyszecki developed the seven-field colorimeter with which an observer can view with both eyes one or more of seven hexagonal fields, each with separately controllable RGB sources, achieved with filtered light and mixed in an integrating sphere. It was successfully used in many different research projects.

Color matching and color-difference matching The MacAdam color-matching error ellipses of 1942 (1 observer) were extended in 1957 to 12 observers by Brown. In 1971 Wyszecki and Fielder determined color-matching error

ellipsoids for three observers [5]. The latter two investigations demonstrated the considerable variability by observer. The seven-field colorimeter was also used for a novel color-difference matching experiment in which three fields were displayed and the observer had to adjust the third field so that its brightness matched the brightness of preselected colors with equal luminance in two fields and its chromaticity resulted in identical perceived differences between the colors in the triangular arrangement [6].

Heterochromatic brightness matching
Wyszecki and coworkers added important experimental data to the luminance of equally bright-appearing stimuli. Many chromatic stimuli, when compared to achromatic ones of the same luminance or luminous reflectance, appear to be lighter or brighter, to be “glowing,” an effect known as the Helmholtz-Kohlrausch effect [7].

Publications Wyszecki authored or coauthored 86 scientific papers and 3 books. The first book, *Farbsysteme*, was published in Germany in 1960 [8], describing various color order systems. He coauthored together with D. B. Judd the second and third editions of the latter's *Color in Business, Science, and Industry*, the third edition published after the passing of Judd [9]. He was the lead author, together with W. S. Stiles, of the monumental *Color Science: Concepts and Methods*,

Quantitative Data and Formulae, with editions in 1967 and 1982 [10]. The second edition remains in print today as a highly important source of information in the field of color science.

References

1. Robertson, A.R.: Necrology of G. Wyszecki, AIC Newsletter No. 3, (Aug) 18–20 (1986)
2. Wyszecki, G.: Valenzmetrische Untersuchung des Zusammenhanges zwischen normaler und anomaler Trichromasie. *Farbe*. **2**, 39–45 (1953)
3. Wyszecki, G.: Evaluation of metamerism colors. *J. Opt. Soc. Am.* **48**, 451–454 (1958)
4. Stiles, W.S., Wyszecki, G.: Counting metamerism object colors. *J. Opt. Soc. Am.* **52**, 313–319 (1962)
5. Wyszecki, G., Fielder, G.H.: New color-matching ellipses. *J. Opt. Soc. Am.* **61**, 1135–1152 (1971)
6. Wyszecki, G., Fielder, G.H.: Color-difference matches. *J. Opt. Soc. Am.* **61**, 1501–1513 (1971)
7. Wyszecki, G.: Correlate for brightness in terms of CIE chromaticity coordinates and luminous reflectance. *J. Opt. Soc. Am.* **57**, 254–257 (1967)
8. Wyszecki, G.: *Farbsysteme*. Musterschmidt, Göttingen (1960)
9. Judd, D.B., Wyszecki, G.: *Color in Business, Science, and Industry*. Wiley, New York (1975) (2nd ed. 1963, 3d ed. 1975)
10. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York (1982) (1st ed. 1967, 2nd ed. 1982)

Xenon Lamp

Wout van Bommel
van Bommel Lighting Consultant, Nuenen,
The Netherlands

Definition

Lamps that produce light as a result of an electrical discharge, generated between two electrodes, in a high-pressure pure xenon gas, which is contained in a transparent bulb.

Note: So-called xenon gas discharge lamps for automobile lighting are in fact metal halide gas discharge lamps in which xenon gas does not contribute to the radiation but is added solely to facilitate ignition of the lamp. These lamps do not belong to the category of pure xenon lamps and are therefore not described here.

Short-Arc Xenon Lamps (Fig. 1)

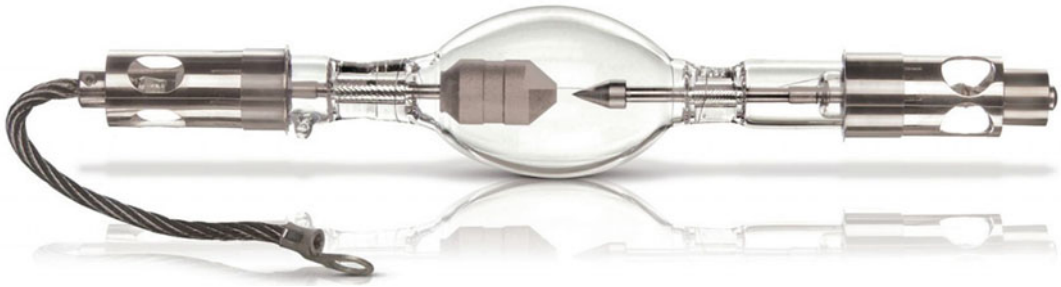
The light-emitting surface of short-arc xenon lamps is a near-point source of very high luminance with excellent and constant color properties. They are developed as an energy-efficient and safe substitute for the awkward and flammable carbon arc lamps used until the 1960s in movie projectors in cinemas. The first movie projection using a short-arc xenon lamp was demonstrated in

1950 in West Berlin by OSRAM. Today's short-arc xenon gas discharge lamps are used for cinema movie projection, film and studio lighting, followspots in theater lighting, searchlights, and lighthouses. Smaller types are used in microscopy, spectroscopy, and optical measurements. The spectrum resembles that of a black body radiator of 6300 K and is very similar to the sun's spectrum.

Long-Arc Lamps (Fig. 2)

Xenon lamps with a larger distance between the electrodes are used in elliptical projector type of sun simulators. They are used for age testing of devices (including solar cells), rapid thermal processing, and material inspection. Versions that deliver short flashes of light are used for photography and industrial applications. They avoid heat build-up in the devices being tested. Their spectrum resembles, just like the short-arc xenon lamp's spectrum, that of a black body radiator of 6300 K, similar to the spectrum of the sun. Where the simulator needs to have a luminance approaching that of the sun's surface, short-arc lamps are employed.

Long-arc xenon lamps were introduced in the late 1950s as an alternative for incandescent lamps in stadium floodlighting. For lighting purposes, these large lamps have since the 1960s been completely superseded by metal halide lamps which have a much higher efficacy.



Xenon Lamp, Fig. 1 Short-arc xenon lamp. (Photograph: Signify)



Xenon Lamp, Fig. 2 Long-arc xenon lamp of the flash type

Working Principle and Construction

Short-arc xenon lamps exist of a single, double-ended, discharge tube of quartz with two electrodes of solid tungsten (Fig. 1). The cathode is cone-shaped with a sharp tip and doped with thorium oxide to facilitate electron emission. Light emission takes place in a very tiny cloud of xenon gas where the electrons leave the cathode tip. The anode is bombarded by the electrons and becomes very hot. It is flat-shaped, large, and of heavy construction to facilitate heat dissipation and increase life. The high-power version lamps need either forced air or water cooling. The very short electrode gap between cathode and anode varies, per type, between some 0.5 and 15 mm. The cone-shaped arc discharge stretches from cathode to anode and is consequently very small and of very high luminance. In vertical burning position, the arc exactly lines up with the electrodes. In horizontal burning position, convection forces may lift up the arc, and a magnet (a current through a coil) is in some types required for a stable, lined up arc position. The very high operating pressure of 10–50 atmospheres requires a cold pressure of some 3–15 atmospheres. During the manufacturing process, therefore, xenon is brought into the discharge tube in frozen form.

Short-arc xenon lamps must be operated on DC power because the very high arc currents require a

different construction of the cathode and anode as described above. The high cold pressure requires high starting voltage peaks up to 50 kV; some types are therefore equipped with an auxiliary electrode to facilitate ignition.

Long-arc xenon lamps have a similar construction as the short-arc versions, but the electrodes are at a larger distance from each other. The light-emitting area of the lamp is longitudinal along the length of the tube.

Properties

System Luminous Efficacy

System luminous efficacies are from 15 lm/W to 50 lm/W. The lower values correspond to lower wattage and lower electrode gap versions.

Lumen Package and Luminance Range

An extensive range of lumen packages is available, varying from some 1000 to more than 1 million lumen (corresponding to some 75–20,000 W). The arc luminance varies, depending on the wattage and electrode gap version, between roughly 20,000 and 500,000 cd/cm^2 (corresponding arc gap, 15–0.5 mm). The higher luminance value is the highest of all electric lamp types and is higher than the luminance of the surface of the sun of approximately 150,000 cd/cm^2 .

Color Characteristics

Contrary to most gas discharge lamps, the spectrum of a short-arc xenon lamp approaches that of a near-continuous spectrum and is practically independent of lamp type, wattage, dimming state, and operating hours. In the visible range, it resembles a black body radiator of 6300 K. The color rendering index is better than 90.

Outside the visible area, the lamps also produce ultraviolet radiation in the region down to 150 nm. The result is that ozone is produced during operation (ultraviolet radiation reacts with oxygen in the air resulting in ozone). In most applications, the oxidizing effect and smell of ozone are unwanted. To avoid these adverse effects, some lamp versions employ doped or coated quartz material that absorbs ultraviolet radiation with wavelengths smaller than 240 nm.

Lamp Life

Average lamp life varies with type from 500 h to 2500 h.

Run-Up and Reignition

Full light output is obtained almost instantaneously. Hot reignition is also immediate.

Handling

The cold pressure of 2–15 atmospheres requires precautions in lamp handling in order to prevent explosion. Protective coverings are supplied with the lamp for safe transportation and handling.

Cross-References

- [Daylighting](#)
- [Metal Halide Lamp](#)

X-Linked Inherited Color Vision Deficiency

- [Deuteranopia](#)
- [Protanopia](#)

XYZ

- [CIE Tristimulus Values](#)